



Universidade de Aveiro Departamento de Biologia
2011

**Conceição Filipa da
Silva Marques**

**Estudo de Polimorfismos Mitocondriais numa
Amostra da População de Goa**



**Conceição Filipa da
Silva Marques**

**Estudo de Polimorfismos Mitocondriais numa
Amostra da População de Goa**

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Aplicada: Biologia Molecular e Celular, realizada sob a orientação científica do Doutor Luís Manuel Souto de Miranda, Professor Auxiliar Convidado do Departamento de Biologia da Universidade de Aveiro.

o júri

Presidente

Doutor João António de Almeida Serôdio,

Professor Auxiliar, Departamento de Biologia da Universidade de Aveiro

Doutor Francisco Manuel Andrade Corte Real Gonçalves

(Arguente Principal), Professor Associado c/ Agregação, Faculdade de Medicina, Universidade de Coimbra, Pólo I – Edifício Central, Rua Larga, 3004-504 Coimbra

Doutor Luís Manuel Souto de Miranda

(Orientador), Professor Auxiliar Convidado, Departamento de Biologia da Universidade de Aveiro

Agradecimentos

Este trabalho é dedicado a todos aqueles que de alguma forma contribuíram para a sua realização, aos quais não posso deixar de agradecer pois sem a colaboração de todos tal não teria sido possível.

Agradeço ao meu Orientador Doutor Luís Souto de Miranda pelos ensinamentos, acompanhamento e disponibilidade fundamentais para o desenvolvimento deste trabalho.

Agradeço às minhas colegas de laboratório, Filipa Tavares, Helena Moreira, Sibylle Gomes e Tânia Barros por todo o incentivo, apoio e contributo durante todo o percurso de execução deste trabalho.

Agradeço à minha família, em especial à minha mãe Maria, ao meu pai Manuel e aos meus irmãos Manuel e Ricardo, sem a ajuda e envolvimento de todos este trabalho não existiria.

Agradeço à minha amiga e irmã afectiva Miriam Mano por todo o carinho, a atenção e o acompanhamento, não imagino os meus dias sem a sua presença.

Agradeço aos meus amigos, a todos aqueles que fizeram parte da minha vida pessoal e académica contribuindo para o meu desenvolvimento enquanto pessoa.

A TODOS O MEU MUITO OBRIGADA.

palavras-chave

Índia, Goa, PCR, ADN mitocondrial, degradação do ADN, haplogrupos mitocondriais.

Resumo

A obtenção do perfil genético em amostras degradadas é frequentemente realizada a partir da análise da informação presente no genoma mitocondrial das amostras. Tal, deve-se às características únicas associadas à molécula do ADNmt, que promovem a resistência a processos de degradação naturais aos quais o ADN está exposto. Contudo esta resistência não impede totalmente a degradação das moléculas do ADNmt, tornando necessário a adequação dos métodos de análise a este tipo de amostras.

A multiplicidade presente a vários níveis na Índia propicia o desenvolvimento de diversos estudos genéticos populacionais que procuram estabelecer se esta diversidade se reflecte no *pool* genético das populações indianas actuais. Neste estudo analisaram-se 135 amostras de habitantes do estado indiano de Goa, através da quantificação por fluorimetria e da PCR em tempo real quantitativa, da amplificação *singleplex* e *multiplex* por PCR, da análise de fragmentos em gel de agarose e da sequenciação automática do ADNmt.

Os resultados confirmaram que as metodologias utilizadas neste estudo são adequadas para análise de amostras degradadas, por isso também indicadas para análise do ADN no âmbito forense.

A qualidade das amostras do ADN revelou-se não ser apropriada para um estudo no âmbito da genética populacional, mas a elevada frequência observada de um haplogrupo caracteristicamente europeu, permite prever uma possível influência de linhagens europeias no genoma mitocondrial da população de Goa.

keywords

Índia, Goa, PCR, mitochondrial DNA, DNA degradation, mitochondrial haplogroups.

abstract

Obtaining the genetic profile of degraded samples is often performed through the analysis of the information present in the mitochondrial genome of the samples. This is due to the unique features associated with the mtDNA molecule, which promotes resistance to natural degradation processes to which the DNA is exposed. However, this resistance does not totally prevent the degradation of mtDNA molecules, which makes it necessary to search for suitable analytic methods for this type of samples.

The multiplicity present at various levels in India promotes the development of several population genetic studies seeking to establish whether this diversity is reflected in the gene pool of the Indian population nowadays.

In this study we analyzed 135 samples of inhabitants of the Indian state of Goa, through the quantification by fluorimetry and real-time quantitative PCR, single and multiplex amplification by PCR, analysis of fragments in agarose gel and automatic sequencing of mtDNA.

The results confirmed that the methodologies used in this study were suitable for degraded DNA samples. Therefore, also indicated for DNA analysis in the forensic field.

The quality of the DNA samples turn out not to be appropriate for a population genetic study, but the high frequency of an European haplogroup observed allows to predict the possible influence of European lineages on the mitochondrial genome of the Goan population.

Índice

Índice de Figuras	ix
Índice de Tabelas	xi
Lista de Abreviaturas	xii
I – Introdução.....	13
1. Genética Humana: passado, presente e futuro	15
2. A molécula do ADN	18
3. Genoma mitocondrial	21
4. Análise do ADNmt humano aplicada à genética populacional e forense	25
5. Amostras degradadas	30
6. A influência portuguesa no estado indiano de Goa	33
6.1.Caracterização geográfica da Índia e Goa.....	33
6.2.Caracterização da população da Índia.....	34
6.3.Estrutura genética da população da Índia.....	35
6.4.História da presença portuguesa em Goa	36
6.5.Caracterização demográfica de Goa	38
7. Objectivos	39
II – Material e Métodos	41
1. Amostras analisadas	43
2. Extracção do ADN.....	43
3. Quantificação do ADN.....	44
4. Amplificação do ADN por PCR.....	47
5. Electroforese em gel de agarose.....	50
6. Purificação do ADN	51
7. Sequenciação automática	51
8. Análise informática de sequências do ADN	52
9. Afiliação em haplogrupos	52
III – Resultados.....	53
1. Quantificação das amostras	55
2. Amplificação e sequenciação da região não codificante.....	60
2.1. Amplificação de todas as amostras de Goa	60
2.2. Amplificação de amostras com extracção recente.....	65
2.3. Sequenciação de amostras teste	70
2.4. Amplificação e sequenciação de fragmentos curtos das amostras teste	73

3. Afiliação em haplogrupos	75
IV – Discussão.....	77
1. Influência de inibidores da PCR	79
2. Presença de contaminação em amostras degradadas	80
3. Determinação da degradação do ADN através da quantificação.....	82
4. Aplicação da estratégia multiplex	83
5. Diversidade genética da população de Goa	86
V – Conclusão	89
VI – Referências Bibliográficas	93
Anexos	101
Anexo 1: Alinhamento das sequências obtidas por sequenciação automática das amostras de Goa	103

Índice de Figuras

Figura 1: Representação da localização celular e estrutura dos genomas nuclear e mitocondrial	18
Figura 2: Representação das subunidades estruturais do ADN	19
Figura 3: Representação da constituição e estrutura da molécula do ADN	19
Figura 4: Diagrama do genoma mitocondrial humano	22
Figura 5: Filogenia dos principais haplogrupos mitocondriais, representativos dos diferentes continentes	28
Figura 6: Representação esquemática dos principais processos de degradação do ADN	30
Figura 7: Mapas da Índia e de Goa. Estão representadas as fronteiras terrestres e marítimas e divisão administrativa	34
Figura 8: Representação esquemática da estratégia A de amplificação para a análise da região não codificante do ADNmt	48
Figura 9: Representação de um gel de agarose com ilustração de fragmentos amplificados de diferentes tamanhos	50
Figura 10: Resultados da amplificação por PCR em tempo real de uma amostra de Goa com diferentes diluições, com a utilização do <i>kit</i> comercial com o fluorocromo <i>SybrGreen</i>	56
Figura 11: Resultados da amplificação por PCR em tempo real de uma amostra de Goa com diferentes diluições, com a utilização do <i>kit</i> comercial com o fluorocromo <i>EvaGreen</i>	57
Figura 12: Resultados da quantificação do fragmento de 164 pb inserido na região HVRI do ADNmt	58
Figura 13: Fotografias de 16 géis de agarose 5% de reacções com o <i>multiplex II</i>	62

Figura 14: Fotografias de 7 géis de agarose 5% de reacções com o <i>multiplex</i> I	64
Figura 15: Fotografias de 4 géis de agarose 5% de reacções com <i>multiplex</i> II	66
Figura 16: Fotografias de 3 géis de agarose a 5% de reacções com <i>multiplex</i> I	67
Figura 17: Exemplo de um electroforetograma de boa qualidade obtido por sequenciação de um fragmento do ADNmt de uma amostra de Goa.	71
Figura 18: Exemplo de um electroforetograma de má qualidade obtido por sequenciação de um fragmento do ADNmt de uma amostra de Goa.	72
Figura 19: Fotografias de 4 géis de agarose 5% de reacções <i>singleplex</i>	74

Índice de Tabelas

Tabela 1 – <i>Primers</i> utilizados para a amplificação do ADN.....	47
Tabela 2 – Conjugação dos <i>primers</i> utilizados para amplificação na estratégia A	48
Tabela 3 – Conjugação dos <i>primers</i> utilizados na estratégia B.....	49
Tabela 4 – Resultados da quantificação do ADN utilizando o fluorímetro	55
Tabela 5 – Resultados da quantificação do fragmento de 164 pb inserido na região HVRI do ADNmt.....	59
Tabela 6 – Síntese dos resultados das amplificações das amostras de Goa com o <i>multiplex</i> I e <i>multiplex</i> II.....	65
Tabela 7 – Síntese dos resultados das amplificações das amostras de Goa com o <i>multiplex</i> I e <i>multiplex</i> II.....	68
Tabela 8 – Comparação geral dos resultados da amplificação das amostras de Goa com o <i>multiplex</i> I e <i>multiplex</i> II com diferentes intervalos decorridos entre o processo de extracção e a análise das amostras extraídas	69
Tabela 9 – Comparação entre amostras com intervalo de extracção/amplificação superior a 1 ano e inferior a 6 meses.....	70
Tabela 10 – Síntese dos resultados de sequenciação das amostras de Goa	73
Tabela 11 – Determinação do haplogrupo de cada amostra através da associação dos respectivos haplótipos aos haplogrupos, com a utilização de duas ferramentas, PhyloTree e mtManager.	75

Lista de Abreviaturas

A	Adenina
ADN	Ácido desoxirribonucleico
ADNmt	ADN mitocondrial
ADNn	ADN nuclear
ARNr	ARN ribossomal
ARNt	ARN de transferência
C	Citosina
Cq	Ciclo de quantificação
CRS	Sequência referência de Cambridge
CRSr	Sequência referência de Cambridge revista
G	Guanina
HVR	Região hipervariável
LCN	<i>Low copy number</i>
pb	Pares de bases
PCR	<i>Polymerase Chain Reaction</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>
rt-qPCR	PCR em tempo real quantitativa
SNP	<i>Single Nucleotide Polymorphism</i>
STR	<i>Short Tandem Repeat</i>
T	Timina
VNTR	<i>Variable Number Tandem Repeat</i>

I – Introdução

(Página intencionalmente deixada em branco)

1. Genética Humana: passado, presente e futuro

O termo genética foi proposto pelo biólogo inglês William Bateson em 1906 para definir o novo ramo da biologia que estava a começar a expandir-se e a ganhar importância dentro da biologia. Este ramo da biologia que tem por objectivo o estudo das regras que regem a hereditariedade e a variação de caracteres entre organismos, suscita grande interesse por parte de toda a comunidade, estando descrito na literatura da antiga Grécia, que já nesse período, a forma como determinadas características físicas humanas eram transmitidas e herdadas, era alvo de reflexão (Passarge, 2007; Stubbe, 1972; Watson & Berry, 2004). Desde então, o estudo da diversidade genética presente no ser humano tem merecido especial atenção, sofrendo uma impressionante evolução, tornando a genética uma ciência multidisciplinar, que marca presença em outras áreas.

A descoberta dos grupos sanguíneos ABO por Landsteiner em 1900 deu início ao estudo da variação genética humana, mas, foi a descrição da variação das frequências destes grupos entre soldados de diferentes países, por Hirszfeld & Hirszfeld em 1919, que marcou o princípio da utilização de marcadores polimórficos em estudos de populações humanas (Hirszfeld & Hirszfeld, 1919; Landsteiner, 1900). Nas décadas que se seguiram, foram descobertos outros polimorfismos sanguíneos, que permitiram a utilização intensiva de métodos imunológicos para a análise de marcadores serológicos (Cavalli-Sforza & Feldman, 2003; Papiha *et al.*, 1999).

Na transição para a segunda metade do século XX, as metodologias utilizadas começaram a sofrer alterações. Os métodos imunológicos foram perdendo a sua relevância, à medida que métodos bioquímicos, moleculares e citológicos começavam a destacar-se como importantes ferramentas de estudo de polimorfismos genéticos. A electroforese foi utilizada por Pauling e colegas em 1949, para a análise de variações presentes na estrutura primária das proteínas, encontrando-se variantes genéticas com utilidade para estudos populacionais (Cavalli-Sforza & Feldman, 2003; Pauling *et al.*, 1949).

Em 1953 Watson e Crick propuseram um modelo da estrutura tridimensional da molécula do ADN, assinalando o início da genética molecular e impulsionando o seu desenvolvimento (Watson & Crick, 1953).

Na década de 80, um novo método, denominado PCR, veio revolucionar o estudo da variação genética. A sua importante contribuição para estudos de ácidos nucleicos garantiu ao seu autor, Kary Mullis, a atribuição do prémio Nobel em 1993. O grande potencial deste método residia na possibilidade de obtenção, de uma grande

quantidade do ADN, a partir de amostras limitadas, isto, recorrendo à multiplicação cíclica das moléculas do ADN (Bertino & Bertino, 2008; Mullis *et al.*, 1986; Mullis & Faloona, 1987).

Nas últimas décadas o desenvolvimento de marcadores moleculares, como SNPs, RFLPs, VNTRs, STRs, mitocondriais e do cromossoma Y, veio facilitar a utilização de métodos moleculares (Speicher *et al.*, 2010).

O desenvolvimento de técnicas de sequenciação veio permitir a determinação da sequência de nucleótidos do ADN, que se estabeleceu como uma prática comum em qualquer laboratório de genética. O aumento da eficiência desta técnica, com a sua automatização, marcou o início de uma nova etapa no conhecimento da variação genética humana, com a determinação de sequências do genoma humano. As primeiras sequências, publicadas por dois grupos distintos em 2001, foram resultado da análise de sequências de vários indivíduos (Lander *et al.*, 2001; Venter *et al.*, 2001). Surgindo, apenas em 2007, a primeira sequência individual, de J. Craig Venter, logo seguida pela publicação da sequência individual de James D. Watson, sucedidas pela publicação de duas sequências anónimas de diferentes origens, uma Asiática e outra Africana (Bentley *et al.*, 2008; Levy *et al.*, 2007; Wang *et al.*, 2008; Wheeler *et al.*, 2008). Desde então, várias sequências individuais completas foram sequenciadas, estando em curso a sequenciação de milhares de sequências (Human genome at ten: The sequence explosion, 2010; Human genome: Genomes by the thousand, 2010).

O progresso colossal da genética humana, principalmente nas últimas décadas, despoletou a sua inclusão em outras áreas da biologia e medicina. Actualmente é notória a importância da genética em diversas áreas da investigação ou medicina, acreditando-se mesmo, que partes da genética serão absorvidas por elas. Por outro lado, é também visível a diversificação da genética em si, aparecendo diversas ramificações (Speicher *et al.*, 2010). Uma das ramificações da genética com maior crescimento nos últimos anos é a genética clínica. Este crescimento deve-se sobretudo à possibilidade de diagnóstico e prognóstico de doenças genéticas, através da análise das variações alélicas dos indivíduos, acreditando-se que futuramente, a identificação de pessoas com elevado risco para certas doenças e a personalização dos tratamentos, promova um crescimento ainda maior da genética clínica (Mills *et al.*, 2011; Speicher *et al.*, 2010). Outra ramificação da genética que tem emergido nos últimos anos é sem dúvida, a genética populacional, sendo representativo do seu crescimento o aumento notório do conhecimento sobre a distribuição da variabilidade genética. Este conhecimento é de grande relevância, uma vez que, para além das implicações éticas e sociais que acarreta, também possui

diversas aplicações, quer ao nível de estudos evolutivos e da origem do ser humano, da distribuição de doenças genéticas, ou mesmo na área forense (Just *et al.*, 2011; McEvoy *et al.*, 2011; Shukla, 2009; Warby *et al.*, 2011; Zhirkova *et al.*, 2011).

O contributo da genética humana para um melhor conhecimento do ser humano, fez recair sobre si enormes expectativas, sendo tida actualmente como uma ciência com um enorme potencial de progressão e desenvolvimento.

2. A molécula do ADN

A célula é a unidade estrutural e funcional básica de todos os seres vivos. De acordo com a sua organização interna, as células podem ser divididas em eucarióticas e procarióticas. As primeiras são maiores e mais complexas, constituídas por várias estruturas citoplasmáticas e organelos, enquanto as segundas possuem menor tamanho e complexidade. Contudo, a principal característica que as distingue, é a presença/ausência de um compartimento específico de armazenamento do património genético delimitado por membranas, o núcleo. É no núcleo, presente apenas em células eucarióticas, que se encontra a maior porção do ADN celular, sendo que, mitocôndrias e cloroplastos possuem também o seu próprio genoma (figura 1) (Azevedo, 2005).

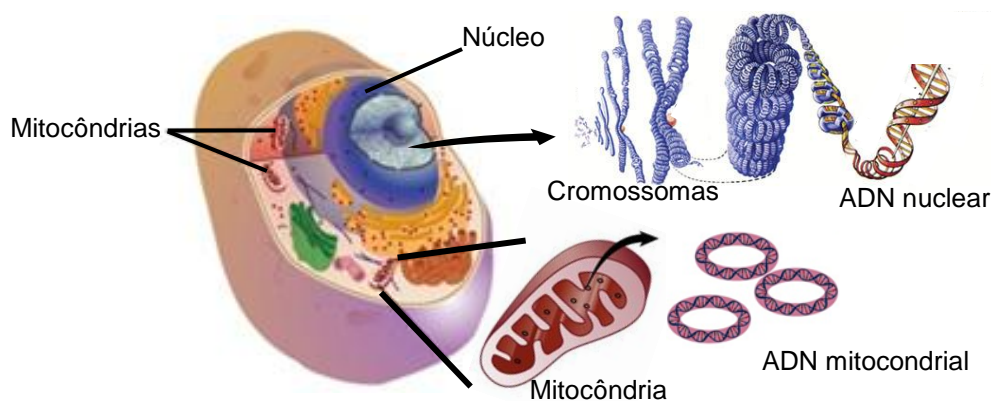


Figura 1: Representação da localização celular e estrutura dos genomas nuclear e mitocondrial (Adaptado de <http://www.research.psu.edu/arp/research-reviews/ultra-deep-sequencing-of-mouse-mitochondrial-dna-mutational-patterns-and-their-origins> e http://www.harunyahya.com/books/darwinism/if_darwin_had_known/if_darwin_had_known05.php, 2011.11.09).

A molécula do ADN é talvez a molécula mais importante das células vivas, é nela que se encontra a informação necessária e indispensável para a sobrevivência das células e do organismo. A sua composição molecular consiste em unidades monoméricas designadas por nucleótidos, constituídos por sua vez por 3 componentes: um açúcar heterocíclico de cinco carbonos, a desoxirribose; uma base heterocíclica azotada derivada de purinas como a adenina ou a guanina, ou derivada de pirimidinas como a citosina ou a timina, ligada ao carbono 1' da desoxirribose por uma ligação glicosídica; e por um grupo fosfato ligado ao carbono 5' da desoxirribose por uma ligação fosfodiéster, é o fosfato que confere a carga negativa à molécula do ADN (figura 2). Os nucleótidos que constituem a molécula são ligados covalentemente por uma ligação fosfodiéster que une o fosfato 5' de um nucleótido ao grupo hidroxilo 3' do nucleótido adjacente, formando longas cadeias do ADN (Butler, 2009; Nelson *et al.*, 2008).

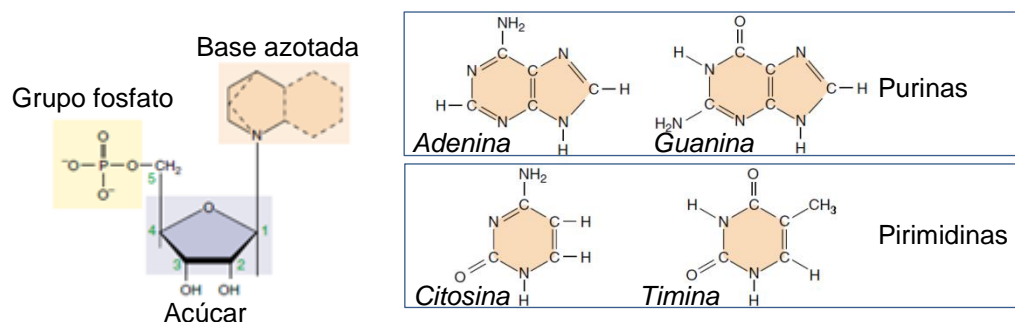


Figura 2: Representação das subunidades estruturais do ADN. À esquerda a estrutura de um nucleótido com as 3 subunidades: um açúcar de 5 carbonos, uma base azotada e um grupo fosfato e à direita as 4 bases azotadas do ADN (*Adaptado de Raven & Johnson, 1999*).

Estruturalmente, a molécula do ADN é composta por duas cadeias polinucleotídicas complementares, enroladas à volta uma da outra, com orientações 5'→3' antiparalelas, formando uma dupla hélice. Na periferia da molécula encontram-se as estruturas desoxirribose-fosfato, sendo que o interior da dupla hélice é ocupado pelos pares de bases purina-pirimidina de ambas as cadeias. Os pares de bases das cadeias complementares ligam-se entre si através de pontes de hidrogénio, sendo que, naturalmente a citosina liga-se à guanina por uma ligação tripla e a adenina à timina por uma ligação dupla (figura 3) (Butler, 2009).

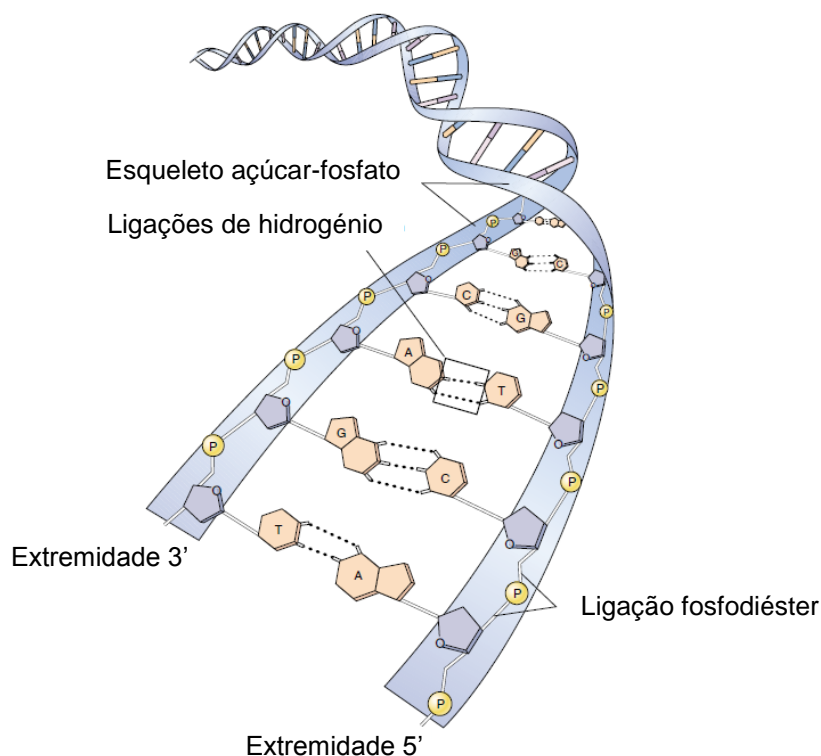


Figura 3: Representação da constituição e estrutura da molécula do ADN (*Adaptado de Lodish, 2008*).

Para além de moléculas do ADN lineares, é possível encontrar também moléculas do ADN circulares, no caso do ser humano presentes nas mitocôndrias. Neste caso, a molécula do ADN tem uma estrutura fechada sem extremidades livres (Lodish, 2008). A percentagem de informação genética total contida em cada um dos genomas do ser humano é bastante diferente, o genoma nuclear constituído por aproximadamente 3.2 mil milhões de pares de bases e representa 99,9995%, com os restantes 0,0005% a serem representados pelo genoma mitocondrial com aproximadamente 16.569 pb (Butler, 2009).

Quando as moléculas do ADN são expostas a determinadas condições, como temperatura elevada ou desnaturantes químicos, ocorre a ruptura das ligações de hidrogénio entre os pares de bases, o que consequentemente provoca a separação das cadeias que formam a dupla hélice, num processo denominado de desnaturação. Este processo é reversível, quando as condições são restabelecidas, as cadeias separadas reúnem-se, formando novamente a dupla hélice. É esta capacidade de ruptura e restituição das ligações entre as duas cadeias da dupla hélice que permite a replicação do ADN e a sua transmissão à descendência, fazendo do ADN a base da hereditariedade (Butler, 2009).

3. Genoma mitocondrial

As mitocôndrias são organelos celulares presentes na grande maioria das células eucarióticas, onde têm como principal função a produção da energia necessária ao funcionamento celular. No caso das células de mamíferos, o número de mitocôndrias por célula apresenta variações, variando normalmente entre 80 e 680 mitocôndrias por célula (Robin & Wong, 1988).

Estes organelos têm-se revelado de extrema importância em vários estudos genéticos, isto devido à presença de um genoma próprio, o ADNmt, que os torna únicos. Geralmente, as mitocôndrias apresentam múltiplas cópias do seu genoma, entre 2 a 8 cópias, fazendo com que cada célula possa possuir milhares de moléculas do ADNmt (Legros *et al.*, 2004; Robin & Wong, 1988).

O genoma mitocondrial humano é uma molécula circular do ADN em cadeia dupla, com aproximadamente 16.569 pb, cujo tamanho varia devido à possibilidade de existência de pequenas mutações. A sua organização é bastante compacta, possui uma maior porção, de cerca de 15.469 pb com função codificante, que codifica 13 subunidades do sistema de fosforilação oxidativa, 2 ARNr, e 22 ARNt (Anderson *et al.*, 1981). Os restantes nucleótidos, cerca de 1.100 pb, são não codificantes, desempenhando funções reguladoras, sendo por isso esta região denominada de região não codificante. As duas cadeias que formam a molécula do ADNmt possuem diferentes distribuições nucleotídicas, numa delas as guaninas, bases de maior peso molecular, encontram-se em maior proporção, sendo por isso esta cadeia denominada de cadeia pesada (H - *Heavy*), designando-se a outra cadeia, rica em citosinas, por cadeia leve (L - *Light*). A replicação da cadeia pesada tem início na origem de replicação presente na região não codificante do ADNmt, promovendo a formação de uma estrutura denominada de *displacement loop* ou D-loop, sendo por isso esta região denominada também de região controlo ou região D-loop (Shadel & Clayton, 1997).

Na região não codificante do ADNmt, identificaram-se 3 segmentos de elevada variabilidade, sendo por isso designados de regiões hipervariáveis, nomeadamente, a região HVRI, localizada entre as posições 16.024 e 16.365 e constituída por 342 pb; a região HVRII, localizada entre as posições 73 e 340 e constituída por 268 pb; e a região HVRIII, localizada entre as posições 438 e 574 e constituída por 137 pb (figura 4) (Greenberg *et al.*, 1983; Lutz *et al.*, 1997).

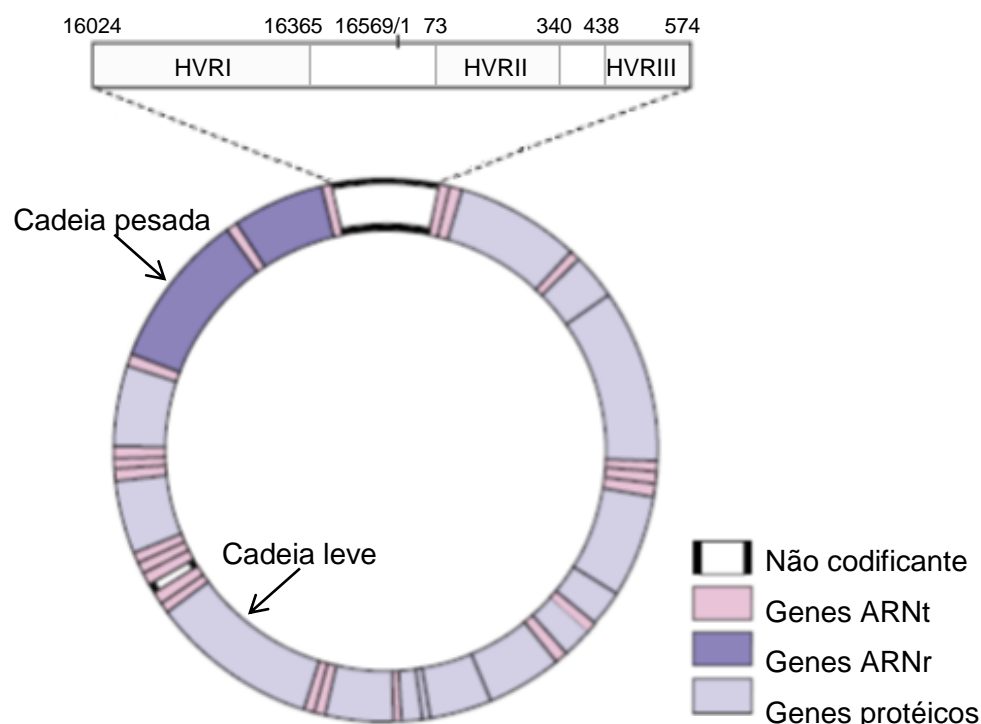


Figura 4: Diagrama do genoma mitocondrial humano. Na região codificante estão representados os 22 genes de ARNt (rosa), os 2 genes de ARNr (roxo) e os 13 genes protéicos (lilás). A região não codificante encontra-se destacada, distinguindo-se as 3 regiões hipervariáveis e os seus limites (*Adaptado de Butler, 2009 e Hagelberg, 2003*).

A sequência completa do ADNmt foi publicada por Anderson e colegas em 1981, sendo denominada de sequência referência de Cambridge (CRS - *Cambridge Reference Sequence*) (Anderson *et al.*, 1981). Anos depois, em 1999, o ADNmt humano foi novamente sequenciado, desta vez por Andrews e colegas, verificando-se a existência de erros em 11 posições nucleotídicas da sequência original, que foram então corrigidos, a esta sequência corrigida deu-se o nome de sequência referência de Cambridge revista (rCRS - *Revised Cambridge Reference Sequence*) (Andrews *et al.*, 1999). Actualmente é a sequência revista a utilizada para comparação em estudos do ADNmt humano.

O ADNmt é uma mais-valia para diversos estudos genéticos, uma vez que é caracterizado por elevada velocidade evolutiva, presença de heteroplasmia, herança uniparental e ausência de recombinação.

O genoma mitocondrial é caracterizado por uma velocidade evolutiva rápida, quando comparada com a velocidade evolutiva do ADNn, a sua é 5 a 10 vezes superior, mas dentro da própria molécula esta velocidade varia, com a região não codificante a apresentar uma velocidade mutacional superior à região codificante, com certas posições nucleotídicas da molécula a comportarem-se como *hotspots* mutacionais (Brown *et al.*,

1979; Howell *et al.*, 2007; Meyer *et al.*, 1999; Pakendorf & Stoneking, 2005; Parsons *et al.*, 1997; Stoneking, 2000). Vários são os factores que contribuem para esta rápida evolução característica do genoma mitocondrial. A elevada velocidade mutacional resultante da ineficiência ou mesmo ausência da função de reparação durante a replicação do ADNmt permite o aparecimento de erros nas moléculas do ADNmt (Clayton *et al.*, 1974; Grossman *et al.*, 1973). Por sua vez, as restrições funcionais e a possibilidade de contornar o aparecimento de mutações num genoma activando outro genoma intacto presente na mesma célula, contribuem para a fixação das mutações (Brown *et al.*, 1979). A elevada sensibilidade do genoma mitocondrial a danos no ADN, provocada, por exemplo, pela ausência de uma protecção proteica (Richter *et al.*, 1988). Como consequência desta rápida evolução, o genoma mitocondrial é altamente polimórfico exibindo uma elevada variabilidade nas populações (Lutz *et al.*, 1998).

A presença de múltiplas cópias do ADNmt por célula e a elevada velocidade mutacional proporcionaram o aparecimento da heteroplasma, ou seja da presença de mais de um tipo do ADNmt num indivíduo, sendo mesmo considerado que todos os indivíduos são heteroplásmicos a algum nível (Bar *et al.*, 2000; Melton, 2004). Vários estudos têm descrito dois tipos de heteroplasma, a heteroplasma pontual e a heteroplasma de comprimento, sendo a segunda a mais frequente. A heteroplasma pontual é caracterizada pela presença de diferentes nucleótidos na mesma posição (Lutz-Bonengel *et al.*, 2008). A heteroplasma de comprimento é geralmente caracterizada pela variação no número de nucleótidos localizados numa extensão homopolimérica (Bar *et al.*, 2000). Na região não codificante do ADNmt estão localizadas quatro segmentos de citosinas repetidas que são altamente susceptíveis à heteroplasma de comprimento, a primeira abrange as posições 16184 bp a 16193 bp, a segunda abrange as posições 303 bp a 309 bp, a terceira abrange as posições 456 bp a 463 bp e a quarta abrange as posições 568 bp a 573 bp (Forster *et al.*, 2010). A heteroplasma é contudo encontrada em baixa frequência, o que indica a existência de um mecanismo celular de regulação dos níveis de heteroplasma, que limita a transmissão de grande variabilidade à descendência (Sondheimer *et al.*, 2011; Wai *et al.*, 2008; White *et al.*, 2008).

No caso dos mamíferos, o ADNmt é de herança exclusivamente materna, devendo-se a ausência do ADN de origem paterna à sua eliminação durante as primeiras divisões embrionárias (Giles *et al.*, 1980; Hutchison *et al.*, 1974; Kaneda *et al.*, 1995; Scarpulla, 2008). O ADNmt de cada indivíduo, à excepção de mutações, é assim igual a toda a linhagem materna, sendo possível relacionar indivíduos de acordo com o seu genoma mitocondrial (Bar *et al.*, 2000). Apesar da herança uniparental no ser humano ser

uma ideia amplamente aceite, o relato da presença do ADNmt paterno no tecido muscular de um indivíduo, correspondente a 90% do seu ADNmt, levantou a possibilidade de herança biparental no ser humano (Schwartz & Vissing, 2002). Mais nenhum caso semelhante foi relatado até a data, pensando-se que este foi um caso isolado de ineficiente eliminação de mitocôndrias de origem paterna, prevalecendo a ideia de herança uniparental materna.

O ADNmt humano é transmitido à descendência sem que haja recombinação entre os genomas paterno e materno (Giles *et al.*, 1980; Hutchison *et al.*, 1974). Durante os primeiros instantes da embriogénese humana, moléculas do ADNmt paterno e materno coexistem sem que haja transferência de informação genética entre ambas, até que mecanismos de eliminação do ADNmt paterno são activados (Kaneda *et al.*, 1995). A descrição de um caso de transferência genética entre ADNmt paterno e materno demonstrou que a recombinação pode ocorrer, contudo, tal como aconteceu com o conceito de herança uniparental, este é um caso atípico, ligado a um estado patológico, não havendo evidências de tais ocorrências em indivíduos saudáveis (Kraytsberg *et al.*, 2004).

Estas características associadas ao ADNmt fazem dele uma ferramenta útil em vários âmbitos de estudo, sendo de realçar a sua utilização em estudos de genética populacional e de genética forense.

4. Análise do ADNmt humano aplicada à genética populacional e forense

Os vários aspectos singulares associados ao ADNmt levaram à sua análise em múltiplos estudos, mas é sem dúvida na genética populacional e na genética forense que a sua utilização se destaca.

Para a genética populacional a importância do ADNmt reside na informação que este transporta sobre as linhagens maternas. Como consequência da herança materna, da elevada velocidade evolutiva e da ausência de recombinação, as mutações que ocorrem no ADNmt acumulam-se nas diferentes linhagens maternas, caracterizando-as. Os polimorfismos do ADNmt presentes nas populações actuais permitem averiguar a estrutura genética, as migrações e cruzamentos das populações do passado, através da transmissão do ADNmt às gerações posteriores. Este processo de diferenciação genética durante as migrações é relativamente rápido tendo ocorrido durante ou após as dispersões. Através da análise do ADNmt é possível perceber onde e quando ocorreram os principais eventos da evolução humana, reconstruindo as dispersões humanas.

O primeiro estudo populacional com análise do ADNmt foi realizado por Brown em 1980, com recurso à digestão enzimática da molécula do ADNmt para análise de RFLPs presentes em 21 amostras de indivíduos de diferentes etnias e origens geográficas (Brown, 1980). Posteriormente o ADNmt começou também a ser alvo de sequenciação, particularmente a região HVRI (342 pb) tem sido o principal alvo de estudos populacionais, por vezes aliada à análise da região HVRII (268 pb), por ser considerada a região de maior variabilidade do genoma mitocondrial (Nunez *et al.*, 2010). Contudo, por vezes é necessário recorrer ao estudo de polimorfismos situados na região codificante do ADNmt, nomeadamente variações numa única posição nucleotídica denominadas SNP (Messina *et al.*, 2010). A combinação destas abordagens permitiu a construção de uma árvore filogenética global, com diversos ramos relacionados entre si, formando grupos de haplótipos mitocondriais, descendentes da mesma molécula do ADNmt ancestral que partilham o mesmo motivo mutacional denominados haplogrupos, que consistem portanto em *clusters* de linhagens intimamente relacionadas (Torroni *et al.*, 2006).

A filogenia do ADNmt tem contribuído para solidificar a ideia de uma origem única do Homem Moderno, tendo já sido demonstrado que a população africana apresenta maior diversidade genética, sendo por isso considerada a população ancestral. Assim, um ancestral comum terá habitado o continente africano há aproximadamente 200.000 anos atrás, migrado e dispersado, povoando progressivamente os restantes

continentes (Cann *et al.*, 1987; Ingman *et al.*, 2000; Macaulay *et al.*, 2005; Vigilant *et al.*, 1991).

As relações filogenéticas entre as diferentes variantes do ADNmt podem ser representadas numa árvore, que reflecte os movimentos migratórios do ser humano. A árvore filogenética do ADNmt é constituída por vários haplogrupos representados pelas letras do alfabeto de A a Z. O ancestral comum materno denominado Eva mitocondrial, é representado pela raiz da árvore, a partir da qual divergem os ramos que constituem os haplogrupos. Mais próximo da raiz está localizado o haplogrupo L, que se divide em dois ramos principais, L0 e L1'2'3'4'5'6, específicos do continente africano e indicativos de uma origem africana. O haplogrupo L0 localiza-se no sul de África enquanto os haplogrupos L1'2'3'4'5'6 estão localizados no leste de África. É a partir do haplogrupo L3 que nasce a maior diversidade genética de populações de origem não-africana, com duas ramificações, os haplogrupos M e N, que representam a colonização euro-asiática, há aproximadamente 63.000 anos, posteriormente à saída de África há cerca de 65.000 anos atrás, surgindo há aproximadamente 60.000 anos o haplogrupo R derivado do haplogrupo N (Forster *et al.*, 2001; Macaulay *et al.*, 2005; Soares *et al.*, 2009; Underhill & Kivisild, 2007; Watson *et al.*, 1997).

Apesar da origem de um ancestral comum em África ser amplamente aceite, as deslocações levadas a cabo pelo ser humano para a colonização da superfície terrestre continuam a gerar controvérsia, não havendo ainda uma ideia clara de como estas ocorreram, mas a conjugação de várias informações genéticas, arqueológicas, históricas e linguísticas, permitem a formulação de várias hipóteses, algumas que vão sucumbindo com as novas descobertas e outras que vão ganhando maior destaque (Knudson & Stojanowski, 2008).

A saída de África das populações ancestrais marcou o início da dispersão e colonização dos outros continentes. A partir de África, as populações terão chegado à Ásia em duas migrações distintas: a primeira seria uma rota localizada a sul, partindo do Corno de África, há cerca de 50.000-100.000 anos, terão atravessado a Arábia Saudita, a costa indiana e o sudeste asiático, chegando à Austrália e à Nova Guiné; a segunda migração terá ocorrido há cerca de 15.000-50.000, pelo norte de África e Vale do Nilo até ao Levante (Cann *et al.*, 1987; Cavalli-Sforza & Feldman, 2003; Macaulay *et al.*, 2005; Quintana-Murci *et al.*, 1999; Stoneking & Delfin, 2010; Xing *et al.*, 2010). A população asiática actual reflecte as migrações dos seus antepassados através dos diversos haplogrupos que a caracterizam, nomeadamente: haplogrupos derivados do haplogrupo M como C, D, E, G e Z; do haplogrupo N como A e Y; e do haplogrupo R como B e F. As

rotas mais longas até a Oceânia também deixaram as suas marcas no genoma mitocondrial, com a população actual a ser caracterizada por alguns haplogrupos bastante restritos: do haplogrupo M os haplogrupos E e Q; do haplogrupo N os haplogrupos O e S; e do haplogrupo R os haplogrupos B e P (Kong *et al.*, 2010; Underhill & Kivisild, 2007; van Oven & Kayser, 2009).

A colonização do continente americano terá ocorrido a partir da dispersão de grupos asiáticos ancestrais originários da Sibéria, que após um isolamento na Beríngia, terão atravessado o estreito de Bering e entrado no novo continente há cerca de 15.000-35.000 anos, tendo então migrando rapidamente para sul e colonizado todo o continente americano (Bonatto & Salzano, 1997; Cavalli-Sforza & Feldman, 2003; Goebel *et al.*, 2008; Schurr & Sherry, 2004; Tamm *et al.*, 2007). Como consequência de ancestrais asiáticos, os haplogrupos mais frequentes no continente americano também fazem parte do grupo de haplogrupos que caracterizam a Ásia: do haplogrupo M os haplogrupos C e D; do haplogrupo N o haplogrupo A; e do haplogrupo R o haplogrupo B (Kong *et al.*, 2010; Underhill & Kivisild, 2007; van Oven & Kayser, 2009). Paralelamente à migração asiática terá ocorrido uma migração euro-asiática para o continente americano, que terá transportado até a América do Norte o haplogrupo X, derivado do haplogrupo N, sem qualquer relação com as populações asiáticas, levantando a possibilidade da população norte americana actual possuir também ancestrais europeus (Brown *et al.*, 1998; Reidla *et al.*, 2003).

A entrada na Europa terá acontecido há cerca de 50.000 anos, a partir do Médio Oriente com a travessia do estreito de Bósforo. A dispersão na Europa terá ocorrido através de duas rotas migratórias: uma pelo litoral oeste mediterrânico até à costa Atlântica de França e até a Grã-Bretanha, e outra atravessando a Europa central até ao norte da Europa (Calafell *et al.*, 1996; Cavalli-Sforza & Feldman, 2003; Richards *et al.*, 1996; Richards *et al.*, 2000). Estas migrações opostas às trajectórias asiáticas reflectem a ausência de haplogrupos provenientes do haplogrupo M, sendo a população europeia com maior frequência caracterizada pelos haplogrupos I, W e X derivados do haplogrupo N; e pelos haplogrupos J, T, HV, H, V, U e K derivados do haplogrupo R (Kong *et al.*, 2010; Underhill & Kivisild, 2007; van Oven & Kayser, 2009).

Na figura 5 estão representados os principais haplogrupos mitocondriais presentes na população mundial.

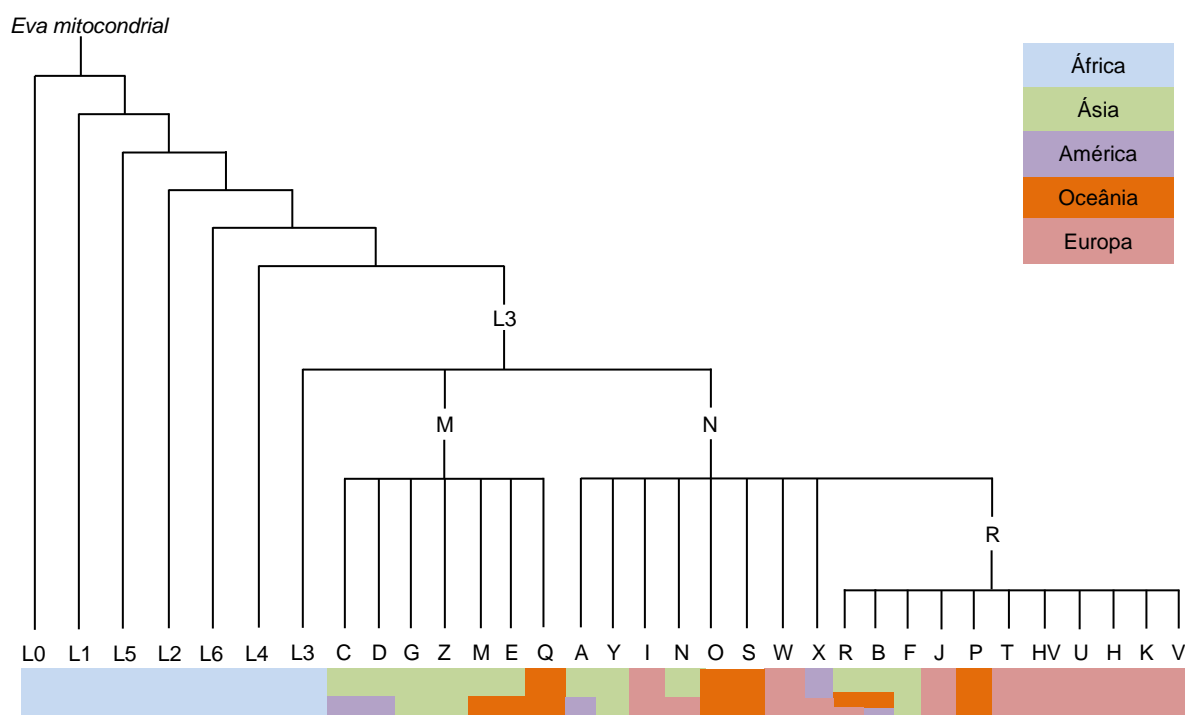


Figura 5: Filogenia dos principais haplogrupos mitocondriais, representativos dos diferentes continentes (Adaptado de Kong *et al.*, 2010; Underhill & Kivisild, 2007 e van Oven & Kayser, 2009).

Nas últimas décadas as populações mundiais têm sido alvo de múltiplos estudos genéticos que utilizam o ADNmt com a aspiração da reconstrução do passado das populações humanas. A complexidade associada aos estudos populacionais dificulta a construção de uma hipótese universalmente aceite sobre o passado do ser humano, permanecendo assim várias hipóteses em aberto, formuladas e fundamentadas nos conhecimentos provenientes do grande número de populações já caracterizadas.

Tal como acontece com a genética populacional, a genética forense tira partido das características que o ADNmt apresenta. O grande número de cópias do ADNmt por célula associado a uma maior resistência à degradação faz do ADNmt o último recurso para obtenção de informação em casos onde a análise do ADNn falha, o que acontece quando o material genético é limitado ou degradado, nomeadamente para a análise de ossos, cabelo, material fixado em blocos de parafina, excrementos e outras amostras expostas a condições adversas (Butler, 2009; Lutz *et al.*, 2000; Mabuchi *et al.*, 2007).

A possibilidade de heteroplasmia do ADNmt num indivíduo pode levar à existência de diferentes tipos do ADNmt em diferentes tecidos ou mesmo em diferentes amostras do mesmo tecido de um indivíduo. Em casos forenses a presença da mesma heteroplasmia em diferentes amostras pode mesmo confirmar a identidade (Ivanov *et al.*, 1996).

A herança uniparental materna e ausência de recombinação levam a que o ADNmt seja idêntico entre todos os indivíduos da mesma linhagem materna, fazendo da sua análise uma ferramenta fundamental em casos forenses, tornando possível a identificação de vítimas ou suspeitos através da comparação do ADNmt de vestígios com o ADNmt de familiares da linhagem materna. Contudo, como resultado da correspondência do ADNmt de toda a linhagem materna e da possibilidade de correspondência com outras linhagens maternas, a tipagem do ADNmt não fornece uma identificação definitiva (Parsons & Coble, 2001).

O reduzido poder discriminatório do ADNmt pode contudo ser atenuado pela construção de bases de dados do ADNmt, através das quais se podem obter informações sobre um determinado haplótipo, como o haplogrupo associado ao haplótipo e qual a frequência desse haplótipo nas populações, o que contribui para um melhor conhecimento individual (Brandon *et al.*, 2005; Ingman & Gyllensten, 2006).

A análise forense do ADNmt é realizada através da sequenciação das regiões HVRI e HVRII, combinadas por vezes com a região HVRIII quando apenas as duas regiões não permitem discriminar indivíduos, ou de toda a região controlo para se obter um maior poder de resolução, mesmo assim alguns indivíduos podem permanecer indistinguíveis, pois muitos polimorfismos são comuns, sendo então necessário recorrer à análise da região codificante (Bini *et al.*, 2003; Brandstatter *et al.*, 2007; Cardoso *et al.*, 2010; Fridman & Gonzalez, 2009; Irwin *et al.*, 2010; Lutz *et al.*, 2000; Mabuchi *et al.*, 2007).

O recurso à análise do ADNmt é documentado em vários casos forenses, onde a identidade de vítimas de catástrofes naturais, de guerras, de actos de terrorismo, de acidentes, de homicídios entre outras, muitas vezes é descoberta ou confirmada devido à comparação dos seus restos mortais com o ADNmt de familiares maternos vivos, servindo também para excluir possíveis vítimas, a informação obtida é muitas vezes fulcral para resolver tais casos (Coble *et al.*, 2009; Gill *et al.*, 1994; Ginther *et al.*, 1992; Grignani *et al.*, 2009; Howard *et al.*, 2011; Ivanov *et al.*, 1996; Just *et al.*, 2011; Sudoyo *et al.*, 2008).

Não há dúvida que o ADNmt tem contribuído bastante para diversos estudos, são as suas características singulares que o tornam útil para uma grande diversidade de aplicações, inclusive na procura de respostas sobre a história das populações humanas na genética populacional e na identificação de indivíduos ligados a cenários forenses.

5. Amostras degradadas

O ADN alvo de análise nem sempre se encontra no melhor estado, como molécula reactiva que é, é sensível a diversos factores que podem alterar as suas propriedades. Principalmente na análise de amostras antigas ou provenientes de contextos forenses é importante ter em conta o passado da amostra e ajustar as metodologias de análise.

Durante o processo de morte celular os processos de reparação do ADN perdem a sua eficiência, não conseguindo contrabalançar os danos provocados por mecanismos enzimáticos, oxidativos, hidrolíticos, entre outros, que vão progressivamente deteriorando a molécula do ADN, auxiliados por factores ambientais como calor, humidade, luz ou fogo (figura 6) (Hall & Ballantyne, 2004; Pääbo *et al.*, 2004).

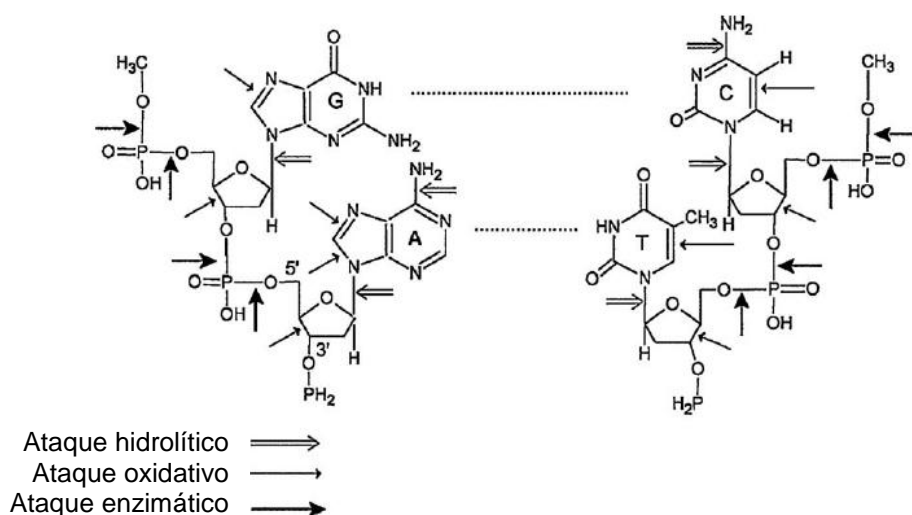


Figura 6: Representação esquemática dos principais processos de degradação do ADN (Adaptado de Alaeddini *et al.*, 2010).

A presença de fluídos celulares ricos em nutrientes na envoltória do ADN é favorável ao aparecimento e desenvolvimento de microrganismos e/ou invertebrados ambientais que se alimentam desses fluídos, reduzindo a quantidade do ADN na amostra ou degradando-o por processos enzimáticos e/ou não-enzimáticos (Alaeddini *et al.*, 2010).

A actividade de enzimas exógenas ou endógenas à célula é um dos processos que mais contribui para a degradação de amostras, uma vez que enzimas como lipases, nucleases e proteases, são capazes de danificar componentes celulares e biomoléculas como o ADN. Os mecanismos de degradação enzimática são variados, desde a eliminação da protecção protéica do ADNn por proteases lisossomais, à clivagem do

ADN por nucleases, sendo a quebra de ligações das cadeias do ADN a consequência final da actividade enzimática (Alaeddini *et al.*, 2010; Pääbo *et al.*, 2004).

O ataque hidrolítico de ligações estabelecidas entre os diferentes constituintes do ADN é também um processo comum de degradação da molécula do ADN. Os danos hidrolíticos resultam principalmente, quer da clivagem de ligações glicosídicas estabelecidas entre a base azotada e a desoxirribose, tendo como consequência a depurinação e a depirimidação, quer da remoção de grupos amino secundários (desaminação) presentes por exemplo na citosina, guanina e adenina, levando à formação de outros compostos (Alaeddini *et al.*, 2010; Gates, 2009; Lamers *et al.*, 2009; Marrone & Ballantyne, 2010; Pääbo *et al.*, 2004).

O ADN pode também sofrer danos oxidativos, causados por espécies reactivas derivadas do oxigénio, como radicais superóxido ($O_2^{\cdot-}$), peróxido de hidrogénio (H_2O_2) e hidroxilo (OH^{\cdot}), resultantes da incidência de radiação ionizante ou da actividade de microrganismos. Os principais alvos do ataque oxidativo são as ligações duplas existentes nos anéis de pirimidinas e purinas, e as ligações na desoxirribose, que resultam em alterações nos respectivos anéis, formação de hidantoínas, eliminação de bases e formação de *crosslinkages* (Alaeddini *et al.*, 2010; Gates, 2009; Keyer & Imlay, 1996; Lamers *et al.*, 2009).

Normalmente, em casos onde as amostras sofreram processos de degradação ou são limitadas, a análise do ADN torna-se inviável, sendo necessário o recurso à análise do ADNmt caracterizado pela sua maior resistência à degradação (Alaeddini *et al.*, 2010). As problemáticas da análise de amostras degradadas manifestam-se principalmente durante a amplificação do ADN através da PCR. As condições óptimas requeridas para a amplificação do ADN alvo não são por vezes fáceis de reunir. Nas amostras degradadas as moléculas do ADN podem estar presentes em número reduzido, denominando-se de amostras LCN, fragmentadas ou modificadas estruturalmente, podendo ainda apresentar ADN contaminante ou inibidores da PCR, o que contribui para a falha da amplificação ou para a ineficiência da reacção (Alaeddini *et al.*, 2010; Gill, 2001; Hall & Ballantyne, 2004). Para contornar as problemáticas associadas à realização da PCR têm sido desenvolvidas diversas estratégias para ajustar esta técnica à análise de amostras degradadas, procedendo-se ao aumento do número de ciclos da PCR, à realização da reacção em *multiplex*, e à redução do tamanho do fragmento do ADN alvo (Berger & Parson, 2009; Gabriel *et al.*, 2001; Grignani *et al.*, 2009).

A exposição de amostras a condições desfavoráveis à preservação do material genético possibilita a sua degradação por diversos processos. Apesar dos desenvolvimentos e adaptações realizados para a análise de amostras degradadas, muitas dificuldades associadas a este tipo de amostras persistem, reduzindo o seu poder informativo, sendo muitas vezes apenas possível, obter perfis genéticos parciais.

6. A influência portuguesa no estado indiano de Goa

6.1. Caracterização geográfica da Índia e Goa

No sul do continente asiático, situada no hemisfério norte, fica localizada a Índia, com o território continental situado entre as latitudes 8°4' e 37°6' norte e as longitudes 68°7' a 97°25' leste, sendo que a extensão de todo o território indiano, chega a 6°45' sul quando as ilhas Nicobar são consideradas (Nag & Sengupta, 1992).

A Índia é o sétimo maior país do mundo, ocupa uma área de 3.287.263 km², com 15.000 km de fronteira terrestre e com uma linha costeira de 7.516 km (Information & Broadcasting, 2010). A Índia faz fronteira com 6 países: o Butão, a China e o Nepal a nordeste; o Paquistão a noroeste; e o Bangladesh e o Mianmar (Birmânia) a leste. Juntamente com o Bangladesh e o Paquistão forma um subcontinente, separado do restante continente asiático pelos Himalaias situados a norte e cordilheiras a oeste e leste. As diferentes linhas costeiras definem a península indiana, a oeste da qual se encontra o Mar Árábico com as ilhas indianas Laquedivas, a leste a baía de Bengala com as ilhas indianas Andamão e Nicobar, e a sul o Oceano Índico onde se localiza o Sri Lanka a aproximadamente 65 km da península (Chary, 2009; Educational & Pletcher, 2010). O território indiano é actualmente dividido em 28 estados, entre os quais o estado de Goa, e 7 territórios de união, tendo como capital a cidade de Nova Delhi (Chandramouli, 2011).

O estado de Goa é composto por uma área continental na costa sudoeste da Índia e a ilha Andegiva situada no Mar Árábico. Actualmente, Goa é o menor estado da Índia com uma área de 3.702 km², situado entre os estados de Maharashtra e Karnataka, limitado a norte pelo rio Tiracol, a sul e a leste pelo estado de Karnataka e a Oeste pelo Mar Árábico (Bhargava, 2006; Educational & Pletcher, 2010).

Goa está dividido administrativamente em 2 distritos, Goa Norte cuja capital é Panaji, também capital do distrito de Goa, e Goa Sul cuja capital é Margão, os distritos Norte e Sul dividem-se por sua vez em 11 concelhos ou *Talukas*. Em Goa é possível encontrar 334 aldeias e 70 cidades, distribuídas por Goa Norte com 194 aldeias e 47 cidades e Goa Sul com 140 aldeias e 23 cidades (figura 7) (Wasnik, 2011).

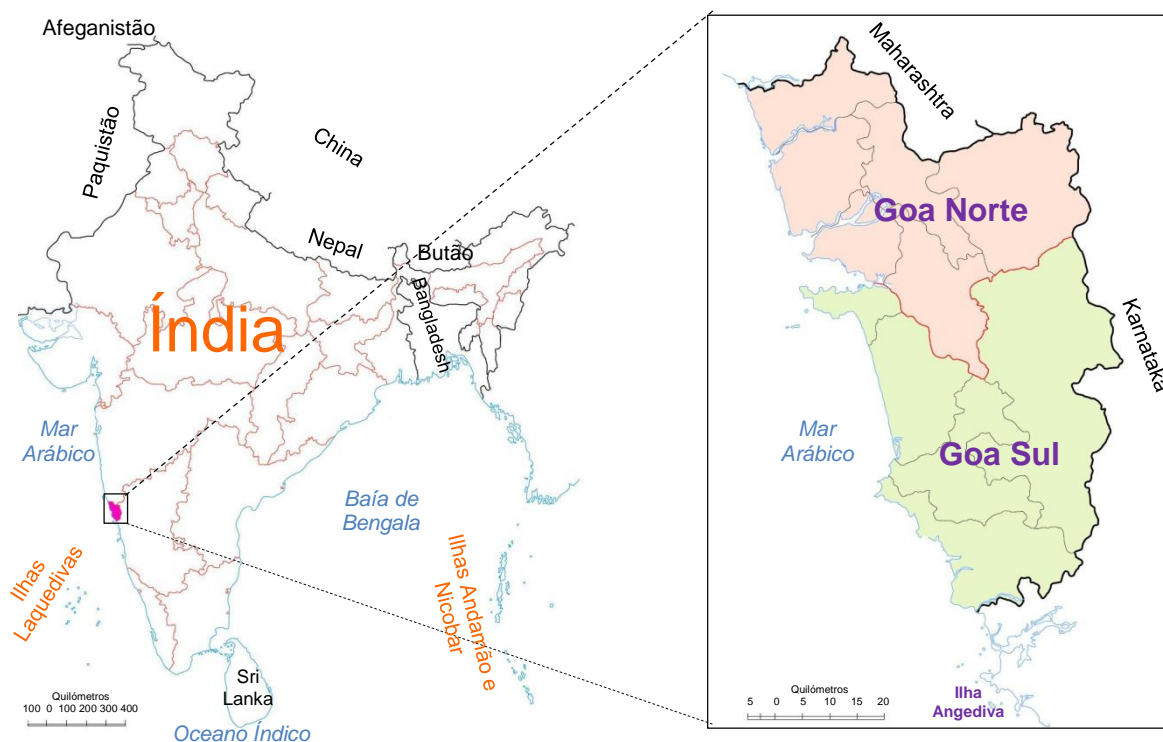


Figura 7: Mapas da Índia e de Goa. Estão representadas as fronteiras terrestres e marítimas e divisão administrativa. Em detalhe encontra-se o estado de Goa, com a divisão em distritos e concelhos (*Adaptado de Wasnik, 2011*).

6.2. Caracterização da população da Índia

A Índia é o segundo país mais populoso do mundo, com uma população de 1.210.193.422 habitantes e uma densidade populacional de 382 habitantes/m², com uma taxa de literacia global de 74,04%, 82,14% no sexo masculino e 65,36% no sexo feminino e uma taxa de crescimento populacional de 17,64% na última década (Chandramouli, 2011). A sociedade indiana apresenta uma estrutura social hierarquizada, com a sua população agrupada em tribos ou castas (Risley & Crooke, 1999). Uma casta pode ser definida como uma selecção de famílias ou grupos de famílias com um nome comum; que alegam descendência comum de um antepassado mitológico, humano ou divino; que reconhecem seguir a mesma vocação hereditária e que formam uma comunidade homogénea (Risley & Crooke, 1999). Por outro lado, uma tribo pode ser definida como uma selecção de famílias ou grupos de famílias com um nome comum; que alegam descendência comum de um antepassado mitológico ou histórico, ou ocasionalmente de um animal; que falam a mesma língua e alegam ocupar uma determinada extensão da Índia (Risley & Crooke, 1999).

As populações indianas nativas têm sido influenciadas por várias populações invasoras, que contribuíram para a diversidade que caracteriza a Índia, denunciada por múltiplos aspectos: a presença de diversos grupos étnicos, provenientes principalmente de grupos indo-arianos e dravidianos; uma multiplicidade linguística, com 22 línguas reconhecidas pela constituição indiana, dominando o hindi falado por 40% da população, e mais de 1.600 dialectos; apesar de não possuir uma religião oficial 80,5 % da população pratica o hinduísmo, com a restante população a praticar diferentes religiões; e diversas culturas (Chary, 2009).

6.3. Estrutura genética da população da Índia

A entrada do Homem Moderno na Índia terá acontecido aquando da migração humana pelas linhas costeiras localizadas a sul dos continentes africano e asiático, ocorrendo posteriormente dispersões pelas várias regiões indianas (Quintana-Murci *et al.*, 1999). A estrutura genética da Índia é resultado de pelo menos 4 componentes: o primeiro é um componente dos ocupantes do Paleolítico Inferior que possivelmente se encontra praticamente extinto actualmente; o segundo consiste na migração de agricultores, que possivelmente falavam linguagens proto-dravidianas, a partir do leste do crescente fértil durante o início do Neolítico; o terceiro, e um dos mais importantes, é a chegada dos arianos, caracterizados pela linguagem indo-europeia, que terão entrado no subcontinente indiano há cerca de 3.500 anos e realizado uma dispersão bem-sucedida para sul e para norte, tendo provocado a deslocação dos dravidianos para sul; O quarto manifesta-se nas populações do nordeste e centro que falam austro-asiático e sino-tibetano, sendo resultado da entrada de populações do leste e nordeste asiático (Cavalli-Sforza *et al.*, 1994). Durante toda a sua história, a Índia comportou-se como um local de chegada e estabelecimento de várias populações oriundas de vários locais, que contribuíram para o seu *pool* genético.

A existência de uma grande diversidade histórica, cultural, linguística, religiosa e geográfica nos habitantes da Índia é propícia ao desenvolvimento de diversos estudos, que procuram perceber se esta diversidade se reflecte no *pool* genético das populações indianas actuais. Os estudos de genética populacional humana que recorrem ao ADNmt debruçam-se sobretudo na análise dos diferentes grupos populacionais dispersos pelas diferentes regiões geográficas indianas.

A população indiana é sobretudo caracterizada por uma maior frequência do haplogrupo asiático M, 60%-64%, seguida pelo haplogrupo europeu U, 14%-21%, com

outros haplogrupos a aparecerem em menor frequência (Baig *et al.*, 2004; Kivisild *et al.*, 1999; Roychoudhury *et al.*, 2000). Na generalidade, apesar do grande número de grupos populacionais existentes, não se notam diferenças significativas nos haplogrupos que os caracterizam, variando contudo a frequência destes haplogrupos. Nas castas, o haplogrupo U é mais frequente do que nos grupos tribais, acontecendo o inverso com o haplogrupo M (Baig *et al.*, 2004; Cordaux *et al.*, 2003; Edwin *et al.*, 2002; Roychoudhury *et al.*, 2000; Thanseem *et al.*, 2006).

A frequência dos haplogrupos entre as diferentes castas da hierarquia apresenta variações, apesar de em todas as castas serem os haplogrupos asiáticos os mais frequentes, nas castas superiores a frequência de haplogrupo asiático M é inferior (61%), quando comparada com as castas médias (65%) e inferiores (71%), acontecendo o contrário com o haplogrupos europeus, com frequência mais elevada nas castas superiores (24%) do que nas castas médias (14%) e castas inferiores (7%) (Bamshad *et al.*, 2001).

Entre os grupos tribais nota-se uma diferença de diversidade, com menor diversidade a sul e superior a norte, com os grupos situados a norte a apresentarem uma frequência menor do haplogrupo M (56%) do que os grupos situados a sul, centro e leste (75%), sendo que no norte e centro há uma maior frequência dos haplogrupos europeus H, JT e W (6-7%), virtualmente ausentes nas tribos do sul, enquanto que a nordeste estão presentes os haplogrupos A e F, frequentes no leste asiático mas ausentes no resto da Índia (Cordaux *et al.*, 2003).

Nos diferentes grupos populacionais indianos estudados é notório o domínio dos haplogrupos M e U, que variam em frequência de acordo com a região geográfica ou com a população estudada.

6.4. História da presença portuguesa em Goa

Já no início do século XVI a Índia tornou-se o destino de várias civilizações europeias: foi local de estadia de portugueses, ingleses, holandeses e franceses, sendo que os portugueses se distinguiam dos outros colonizadores europeus, pelo facto de encorajarem o casamento com mulheres nativas, prática geralmente desencorajada pelos restantes (Cavalli-Sforza *et al.*, 1994). Os portugueses foram os primeiros colonizadores a chegarem à Índia, tendo alcançado a costa ocidental em 1498 por via marítima e a partir daí conquistado vários territórios indianos (Chary, 2009).

Uma das principais conquistas portuguesas na Índia ocorreu em 1510, com a ocupação de Goa, um território localizado na costa sudoeste da Índia, habitado então maioritariamente por hindus e muçulmanos, que se tornaria a base portuguesa na Índia com o estatuto de capital do Estado Português da Índia (Borges & Feldmann, 1997). Nos anos que se seguiram a estadia portuguesa começou a reflectir-se no perfil demográfico da região, com alterações étnicas dos seus habitantes, resultantes da coabitação e interacção de portugueses e indianos, encorajada pelos líderes portugueses com o propósito de aumentar a comunidade portuguesa em Goa, fortalecendo o seu domínio. Nos primeiros anos de ocupação portuguesa, os habitantes portugueses em Goa eram na sua grande maioria do sexo masculino, uma vez que dificuldades financeiras e factores como o longo percurso até Goa, o clima quente e prevalência de doenças em Goa desencorajavam a emigração feminina (Gracias, 1996). A superioridade masculina proporcionou o estabelecimento de relacionamentos casuais entre homens portugueses e mulheres nativas. Na tentativa de legalizar estes relacionamentos, Portugal estabeleceu a *Política dos Casamentos*, que consistia na formação de alianças matrimoniais entre homens portugueses e mulheres nativas, incentivada por recompensas económicas ou obtenção de cargos governamentais, resultando na formação de uma comunidade mista (Borges & Feldmann, 1997; Gracias, 1996). A crescente incorporação da língua e costumes locais na comunidade portuguesa em Goa levou o Governo português a enviar mulheres portuguesas para Goa, procurando manter um estilo de vida característico da sociedade portuguesa. Assim, em 1545 Portugal começou a enviar órfãs, como potenciais noivas para os portugueses já estabelecidos em Goa, conferindo benefícios aos homens que as aceitassem (Fairchilds, 2007).

Estes deslocamentos de portugueses para Goa modificaram a estrutura da comunidade nativa em vários níveis, quer a nível global, introduzindo a cultura portuguesa, através dos costumes, língua, religião e arquitectura, quer a nível mais individual, introduzindo características portuguesas nos descendentes dos vários relacionamentos estabelecidos durante o período de ocupação. Goa esteve sobre influência portuguesa durante aproximadamente 450 anos, interrompida apenas por uma ocupação inglesa temporária em 1809. Logo após a independência da Índia em 1947, os indianos reclamaram Goa, contudo apenas em 1962, como resultado da invasão indiana em 1961, Goa foi novamente integrada na União Indiana, tornando-se um estado Indiano em 1987 (Bhargava, 2006; Educational & Pletcher, 2010).

6.5. Caracterização demográfica de Goa

A população de Goa é constituída por 1.457.723 habitantes, 740.711 do sexo masculino e 717.012 do sexo feminino, com uma densidade populacional de 394 habitantes/m² e um crescimento populacional de 8,17% na última década. A taxa de literacia da população de Goa é de 87,40%, 92,81% no sexo masculino e 81,84% no sexo feminino (Wasnik, 2011). A língua oficial é o concani, falada pela maioria da habitantes, que se distribuem por diversas comunidades, principalmente cristãs e hindus, onde ainda hoje é possível encontrar pessoas com nomes e apelidos portugueses, o que reflecte a permanência da influência portuguesa neste estado indiano (Chandramouli, 2011; Educational & Pletcher, 2010; Wasnik, 2011).

7. Objectivos

No presente trabalho pretende-se analisar o ADNmt de uma amostra de 135 indivíduos da população de Goa, Índia, enquadrando o trabalho no contexto da genética forense e genética populacional e testar metodologias para aferir o grau de degradação de amostras (por exemplo PCR em tempo real).

No âmbito da genética forense pretende-se aplicar estratégias de amplificação em *singleplex* e *multiplex* de fragmentos curtos da região não codificante do ADNmt em amostras degradadas, de forma a obter a sequência completa da região não codificante do ADNmt. Neste contexto, pretende-se caracterizar a população de Goa de acordo com os haplótipos mitocondriais, disponibilizando informação potencialmente útil para investigações forenses.

No âmbito da genética populacional pretende-se caracterizar a diversidade genética mitocondrial da população de Goa através da afiliação dos haplótipos a haplogrupos do ADNmt, contextualizando a amostra da população de Goa na Índia.

Assim, o principal objectivo deste estudo é a caracterização da estrutura genética de um grupo de amostras do ADNmt de habitantes de Goa, recorrendo a estratégias utilizadas em contextos forenses.

(Página intencionalmente deixada em branco)

II – Material e Métodos

(Página intencionalmente deixada em branco)

1. Amostras analisadas

Para a realização deste estudo foram analisadas 135 amostras, 65 do sexo feminino e 70 do sexo masculino, com 98 cedidas por indivíduos com nomes e/ou apelidos portugueses e 37 cedidas por indivíduos com nomes indianos, todas provenientes de habitantes de Goa, Índia. O material biológico utilizado consistiu em sangue, presente sob a forma de mancha em suporte de tecido de algodão. As amostras permaneceram armazenadas à temperatura ambiente de forma não individualizada, durante pelo menos 10 anos, tendo-se procedido à sua separação para envelopes individuais a aproximadamente 4 anos.

2. Extracção do ADN

Para se aumentar a quantidade e pureza do ADN obtido de cada amostra recorreu-se à extracção do ADN através do método modificado de extracção por Chelex (Walsh *et al.*, 1991). Este é um método que utiliza uma resina quelante com grande afinidade para iões metálicos, com capacidade de remoção destes iões presentes em solução. A alcalinidade da solução de Chelex provoca a desnaturação do ADN extraído, obtendo-se no final ADN de cadeia simples ((Butler, 2011)). As principais vantagens deste método são a execução rápida e simples, baixo custo e a ausência de transferência de líquido entre tubos (Goodwin *et al.*, 2011).

A solução de extracção Chelex 5% foi preparada anteriormente ao início do processo de extracção, consistindo em 5 g da resina Chelex 100 (Bio-Rad), 1 mL de solução tampão Tris-EDTA (Sigma), 1 mL de solução de SDS 20% preparada com SDS 99% (Sigma), e aproximadamente 95 mL de água *Mili-Q* autoclavada. O pH da solução de extracção encontrava-se sempre entre 9.00 e 11.00, determinado com o medidor de pH C830 (Consort).

O processo de extracção foi iniciado com a transferência de uma porção de aproximadamente 3 mm x 3 mm (9 mm²) da mancha de sangue no suporte de tecido para um microtubo de 1,5 mL. Para remover possíveis contaminantes e inibidores procedeu-se à lavagem da amostra através da adição de 1 mL de água *Mili-Q* autoclavada, seguida por um período de incubação de 30 minutos à temperatura ambiente. Para a separação dos componentes a amostra foi colocada na centrífuga 1-14 (Sigma) a 14.000 rpm durante 3 minutos, posteriormente o sobrenadante foi removido e

descartado, permanecendo apenas 25 µL no microtubo. Procedeu-se à adição de 175 µL de solução de Chelex 5%, mantida em agitação contínua no agitador magnético 375 *hot plate stirrer* (VWR).

Realizou-se uma nova incubação, desta vez durante 30 minutos a 56°C e 400 rpm no termomisturador *Confort* (Eppendorf), seguida por uma ligeira agitação no vórtex *mini Vortex V* (VWR) durante 5 a 10 segundos, com uma posterior incubação a 99°C durante 8 minutos no termomisturador ocorrendo desnaturação do ADN, disrupção das membranas celulares e destruição das proteínas, realizou-se novamente uma agitação no vórtex durante 5 a 10 segundos. A extracção foi concluída com uma centrifugação durante 3 minutos a 14500 rpm, armazenando-se o material a -20°C até a sua análise.

3. Quantificação do ADN

A quantificação do ADN fornece informações importantes para a realização do trabalho, indicando se a amostra contém a quantidade do ADN necessária para os procedimentos realizados posteriormente.

Foram realizados dois tipos de quantificação do ADN, um com a utilização do fluorímetro *Qubit®* (Invitrogen) para quantificação do ADN extraído, e outro, quantificação simultânea à amplificação do ADN, recorrendo-se à rt-qPCR no termociclador *C1000™* (Bio-Rad) com sistema de detecção *CFX96™* (Bio-Rad).

3.1. Quantificação por fluorímetro

Para a realização da quantificação do ADN no fluorímetro utilizou-se o *kit* comercial *Quant-It™ dsDNA BR Assay Kit* (Invitrogen) seguindo-se as indicações do fornecedor. As soluções necessárias foram utilizadas à temperatura ambiente. Preparou-se a solução *Quant-iT™ working solution* fazendo-se a diluição de 1:200 do reagente *Quant-iT™ dsDNA BR reagent* em tampão *Quant-iT™ dsDNA BR buffer*. A solução *Quant-iT™ working solution* foi dividida pelos *standards* e amostras, com um volume de 190 µL de solução por cada, e adicionaram-se 10 µL de *standard* ou amostra, totalizando-se 200 µL, agitando-se 2 a 3 segundos no vórtex *mini Vortex V* (VWR) e colocados à temperatura ambiente para uma incubação de 2 minutos. Procedeu-se então à calibração do fluorímetro com o *standard 1* e *standard 2* e realizou-se a leitura da concentração das amostras.

3.2. Quantificação por rt-qPCR

Para a realização da quantificação do ADN por rt-qPCR foram testados dois fluorocromos diferentes, *SybrGreen®* (Bio-Rad) e *EvaGreen®* (Bio-Rad).

Para o teste do *kit* comercial com o fluorocromo *SybrGreen* foi utilizada uma amostra de Goa de forma não diluída, diluída 10x e diluída 100x, tendo-se primeiro diluído o ADN e depois preparado a mistura de reacção com: *iQ SybrGreen supermix* (2x) constituído por 100 mM KCl, 40 mM Tris-HCl pH 8.4, 0.4 mM de cada dNTP (dATP, dCTP, dGTP e dTTP), iTaq ADN polimerase 50 unidades/mL, 6 mM MgCl₂, *SybrGreen I*, 20 nM fluoresceína e estabilizadores, com uma concentração final de 1x, *primer* F15989 a 200 nM de concentração final, *primer* R16152 a 200 nM de concentração final e água *Mili-Q* autoclavada para perfazer um volume de mistura de reacção de 19 µL por reacção. A mistura de reacção foi distribuída por tubos de 1,5 mL adicionando-se a diluição do ADN correspondente, à excepção do controlo negativo onde não é adicionado ADN, agitando-se no vórtex e centrifugando-se suavemente. A mistura de reagentes e ADN foi distribuída pela placa (Bio-Rad), com 20 µL de volume total por poço, com duas réplicas por diluição, selando-se a placa com uma película adesiva (Bio-Rad). A reacção foi realizada no termociclador *C1000™* (Bio-Rad) com sistema de detecção *CFX96™* (Bio-Rad) com as seguintes condições: activação enzimática e desnaturação inicial a 95°C durante 10 minutos; 45 ciclos de desnaturação a 95°C durante 15 segundos e hibridização/extensão 48°C durante 1 minuto; e curva de *melting* de 65°C a 95°C, com intervalo de 0,5°C por 5 segundos.

Para o teste com o fluorocromo *EvaGreen* foi utilizada uma amostra de Goa, a mesma utilizada com o teste do *kit* comercial *SybrGreen*, de forma não diluída, diluída 10x e diluída 100x, tendo-se primeiro diluído o ADN e depois preparado a mistura de reacção com: *SsoFast EvaGreen supermix* (2x) constituído por tampão de reacção com dNTPs, *Sso7s-fusion* polimerase, MgCl₂, corante *EvaGreen* e estabilizadores, com uma concentração final de 1x, *primer* F15989 a 200 nM de concentração final, *primer* R16152 a 200 nM de concentração final e água *mili-Q* autoclavada para perfazer um volume de mistura de reacção de 19 µL por reacção. A mistura de reacção foi distribuída por tubos de 1,5 mL adicionando-se a diluição do ADN correspondente, à excepção do controlo negativo onde não é adicionado ADN, agitando-se no vórtex e centrifugando-se suavemente. A mistura de reagentes e ADN foi distribuída pela placa (Bio-Rad), com 20 µL de volume total por poço, com duas réplicas por diluição, selando-se a placa com uma película adesiva (Bio-Rad). A reacção foi realizada no termociclador *C1000™* (Bio-Rad)

com sistema de detecção *CFX96TM* (Bio-Rad) com as seguintes condições: activação enzimática e desnaturação inicial a 98°C durante 2 minutos; 47 ciclos de desnaturação a 98°C durante 5 segundos e hibridização/extensão a 48°C durante 10 segundos; e curva de *melting* de 65°C a 95°C, com intervalo de 0,5°C por 10 segundos.

A quantificação absoluta com recurso à *rt-qPCR* foi então realizada com a utilização do fluorocromo *EvaGreen* (Bio-Rad).

Para a realização da curva padrão foi utilizado o *kit* comercial *SsoFastTM EvaGreen® Control Assay*. Em primeiro lugar procedeu-se à diluição do ADN, fornecido no *kit*, em água *Mili-Q* autoclavada, obtendo-se oito diferentes concentrações: 10⁸ cópias de ADN/μL; 10⁷ cópias de ADN/μL; 10⁶ cópias de ADN/μL; 10⁵ cópias de ADN/μL; 10⁴ cópias de ADN/μL; 10³ cópias de ADN/μL; e 10² cópias de ADN/μL. Seguidamente preparou-se a mistura de reacção, com *SsoFast EvaGreen supermix (2x)* a uma concentração final de 1x, mistura de *primers* 100x com cada *primer* a 50 μM a uma concentração final de cada *primer* de 500 nM e água *Mili-Q* autoclavada, agitando suavemente para misturar todos os componentes. A mistura de reacção foi distribuída por tubos de 1,5 mL adicionando-se a diluição do ADN correspondente, à excepção do controlo negativo onde não é adicionado ADN, agitando-se no vórtex e centrifugando-se suavemente. A mistura de reagentes e ADN foi distribuída pela placa (Bio-Rad), com 20 μL de volume total por poço, com duas réplicas por diluição, selando-se a placa com uma película adesiva (Bio-Rad). A reacção foi realizada no termociclador *C1000TM* (Bio-Rad) com sistema de detecção *CFX96TM* (Bio-Rad) com as seguintes condições: activação enzimática e desnaturação inicial a 95°C durante 30 segundos; 40 ciclos de desnaturação a 95°C durante 1 minuto e hibridização/extensão 48°C durante 5 segundos; e curva de *melting* de 60°C a 95°C, com intervalo de 0,5°C por 5 segundos.

Para quantificação das amostras de Goa foi portanto utilizado o *kit* comercial *SsoFastTM EvaGreen®*. Em primeiro lugar procedeu-se à preparação da mistura de reacção com: *SsoFast EvaGreen supermix (2x)* com uma concentração final de 1x, *primer* F15989 a 200 nM de concentração final, *primer* R16152 a 200 nM de concentração final e água *mili-Q* autoclavada para perfazer um volume de mistura de reacção de 19 μL por reacção. A mistura de reacção foi distribuída pela placa (Bio-Rad) e adicionou-se a diluição do ADN correspondente, à excepção do controlo negativo onde não é adicionado ADN, perfazendo um volume de reacção total de 20 μL por poço, selando-se a placa com uma película adesiva (Bio-Rad). A reacção foi realizada no termociclador *C1000TM* (Bio-Rad) com sistema de detecção *CFX96TM* (Bio-Rad) com as seguintes condições: activação enzimática e desnaturação inicial a 98°C durante 2

minutos; 47 ciclos de desnaturação a 98°C durante 5 segundos e hibridização/extensão a 48°C durante 10 segundos; e curva de *melting* de 65°C a 95°C, com intervalo de 0,5°C por 10 segundos.

4. Amplificação do ADN por PCR

O ADN das amostras extraídas foi amplificado por PCR de acordo com duas estratégias diferentes, estratégia A e estratégia B, tendo sempre como alvo a região não codificante do ADNmt, dividida em vários segmentos. As amplificações foram realizadas no termociclador *iCycler* (Bio-rad) de acordo com as instruções indicadas pelo fornecedor dos *kits* de reacção utilizados, sendo sempre realizada a amplificação de um **controlo negativo**, onde a amostra foi substituída por água *Mili-Q* autoclavada. Para a amplificação do ADN em reacções com mais de um par de *primers*, denominado de *multiplex*, foi utilizado o *kit* comercial *Multiplex PCR Kit* (QIAGEN), enquanto que as reacções com apenas um par de *primers*, denominadas de *singleplex*, foram realizadas com a utilização do *kit* comercial *Taq PCR Core Kit* (QIAGEN). Os *primers* utilizados para a amplificação, tabela 1, foram desenhados e utilizados por Eichmann & Parson, 2008 .

Tabela 1 – *Primers* utilizados para a amplificação do ADN. Com a designação atribuída, alcance na sequência do ADN e a sequência do *primer* (Eichmann & Parson, 2008).

Designação	Alcance	Sequência 5'-3'
F15989	15989-16008	CCCAAAGCTAAGATTCTAAT
R16152	16130-16152	AGGTGGTCAAGTATTTATGGTAC
F16094	16094-16112	TCGTACATTACTGCCAGYC
R16248	16228-16248	GTTGCAGTTGATGTGTGATAG
F16197	16197-16221	CTTACAAGCAAGTACAGCAATCAAC
R16433	16412-16433	TTGTGCGGGATATTGATTTTAC
R16509	16490-16509	AGGAACCAGATGTCTGGATAC
F16450	16450-16470	GCTCCGGGCCCATAA CACTTG
F16533	16533-16551	TAGCCACACGTTCCCTT
R180	156-180	AATATTGAACGTAGGTGCGATAAAT
F109	109-132	GCACCCTATGTCTGCAGTATCTGTC
F317	317-338	CTTCTGGCCACAGCACTTAAAC
R460	440-460	AGGGGAAAATAATGTGTTAGT
F402	402-425	ATCTTTTGGCGGTATGCACTTTTA
R619	595-619	AACATTTTTCAGTGTATTGCTTTGAG
R599	580-599	TTGAGGAGGTAAGCTACATA

Como já referido a amplificação do ADNmt foi realizada com duas estratégias diferentes, a **estratégia A** e a **estratégia B**, de acordo com o tamanho dos amplicões, pretendendo-se complementar a informação obtida por uma das estratégias com a informação obtida pela outra.

Na **estratégia A** os *primers* foram agrupados formando amplicões de tamanhos entre 299 pb e 444 pb, tendo-se procurado a amplificação de toda a região controlo do ADNmt, através de duas reacções *multiplex* em separado. Uma reacção utilizando 3 pares de *primers*, denominada de *multiplex* I, e a outra reacção utilizando 2 pares de *primers*, denominada de *multiplex* II, com *primers* agrupados segundo Berger & Parson, 2009 de modo a formar amplicões que em conjunto cobriam toda a região não codificante, sobrepondo-se entre si nas extremidades (tabela 2 e figura 8).

Tabela 2 – Conjugação dos *primers* utilizados para amplificação na estratégia A. A fundo branco estão representados os fragmentos do ADNmt amplificados conjuntamente no *multiplex* I. A fundo cinzento estão representados os fragmentos do ADN amplificados conjuntamente no *multiplex* II. *Os *primers* R619 e R599 representam uma alternativa um ao outro para utilização com o *primer* F317.

Primer Forward	Primer Reverse	Tamanho do amplicão (pb)	Multiplex
F15989	R16433	444	I
F16197	R16509	312	II
F16450	R180	299	I
F109	R460	351	II
F317	R619*	302	I
	R599*	282	I

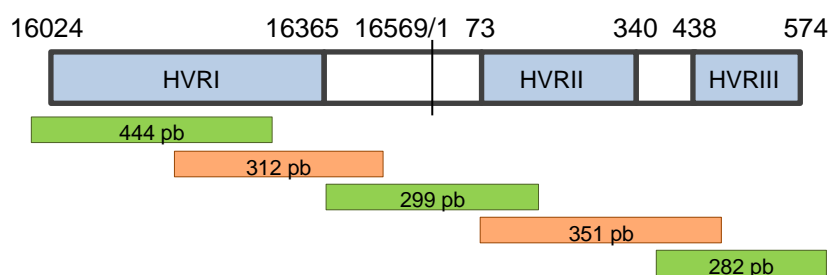


Figura 8: Representação esquemática da estratégia A de amplificação para a análise da região não codificante do ADNmt. A verde estão representados os fragmentos do ADNmt amplificados conjuntamente no *multiplex* I. A laranja estão representados os fragmentos do ADN amplificados conjuntamente no *multiplex* II (Adaptado de Berger & Parson, 2009).

O volume total das reacções sofreu alterações no decurso do estudo, adaptando-se o volume total à quantidade de produto necessária para os procedimentos

realizados posteriormente à amplificação, tendo-se mantido sempre as mesmas proporções de reagentes, com o volume do ADN de 1 µL, ou excepcionalmente 2 µL do ADN diminuindo o volume de água. Para um volume total de 25 µL os componentes das reacções consistiram em 1 µL do ADN extraído ou de água *Mili-Q* autoclavada no controlo negativo e 24 µL de mistura de reacção, constituída por: 12,5 µL de *QIAGEN Multiplex PCR Master Mix* (3 mM MgCl₂); 2,5 µL de mistura de *primers* (200 nM cada *primer*), previamente preparado por diluição dos *primers* em água *Mili-Q* autoclavada; e 9 µL de água *Mili-Q* autoclavada. As reacções foram realizadas no termociclador (Bio-Rad) segundo as seguintes condições: desnaturação inicial durante 15 minutos a 95°C; 39 ciclos de desnaturação durante 30 segundos a 94°C, hibridização durante 90 segundos a 57°C e extensão durante 90 segundos a 72°C; e extensão final durante 10 minutos a 72°C.

Na **estratégia B** os *primers* foram agrupados formando amplicões de tamanhos entre 144 pb e 237 pb, tendo-se procurado obter por amplificação, segmentos não obtidos com a estratégia A, reduzindo o tamanho dos amplicões, complementando-se assim a informação da estratégia A. As reacções foram realizadas utilizando apenas 1 par de *primers* por reacção, combinados segundo Eichmann & Parson, 2008 (tabela 3).

Tabela 3 – Conjugação dos *primers* utilizados na estratégia B.

<i>Primer Forward</i>	<i>Primer Reverse</i>	Tamanho do amplicão (pb)	Temperatura de hibridização (°C)
F15989	R16152	164	48
F16094	R16248	155	52,6
F16533	R180	217	60
F402	R619	218	61

Os componentes das reacções consistiram em 1 µL do ADN extraído, ou excepcionalmente 2 µL do ADN diminuindo o volume de água, e 24 µL de mistura de reacção, constituída por: 2,5 µL de 10× *PCR Buffer* (15 mM MgCl₂); 0,5 µL *dNTP mix* (10 mM de cada em solução); 0,2 µL *Taq DNA polimerase*; 1 µL *primer forward* [10 nm] a concentração final de 160 nM; 1 µL *primer reverse* [10 nm] a concentração final de 160 nM; 18,8 µL de água *Mili-Q* autoclavada. As reacções foram realizadas no termociclador *iCycler* (Bio-Rad) segundo as seguintes condições: desnaturação inicial durante 3 minutos a 94°C; 34 ciclos de desnaturação durante 30 segundos a 94°C, hibridização durante 90 segundos à respectiva temperatura de hibridização de cada *primer forward* e extensão durante 60 segundos a 92°C; e extensão final durante 10 minutos a 72°C.

5. Electroforese em gel de agarose

Os fragmentos do ADN amplificado por PCR foram analisados em gel de agarose a 5%, para se determinar se a amplificação foi bem-sucedida através da visualização do ADN no gel de agarose. Para a preparação do gel, TBE 10x (National Diagnostics) diluído a 1x foi adicionado à agarose *SeaKem® LE Agarose* (Lonza) de forma a obter uma solução de 5% de agarose em TBE, que posteriormente foi aquecida no microondas até ocorrer a dissolução da agarose. O volume de TBE evaporado durante o aquecimento foi repostado. Adicionaram-se 4 µL de corante *GelRed™ Nucleic Acid Gel Stain 10,000x in water* (Biotium), agitando-se manualmente para misturar os componentes deixando-se o gel arrefecer ligeiramente. O gel foi então vertido na moldeira já nivelada e com o pente colocado, ficando a solidificar à temperatura ambiente durante 30 minutos. Para a preparação do marcador molecular utilizaram-se 4 µL de água *Mili-Q*, 1 µL de *GeneRuler™ 100 bp DNA Ladder, 100-1000 bp* (Fermentas) e 1 µL de *6x DNA Loading Dye* (Fermentas). Para a análise do ADN utilizaram-se 5 µL do ADN amplificado por PCR ao qual se adicionou 1 µL de *6x Orange DNA Loading Dye* (Fermentas). Após a solidificação do gel, retirou-se o pente e colocou-se o gel na tina de electroforese ligada ao fornecedor de energia *PowerPac Basic™* (Bio-rad) com TBE 1x, o marcador e as amostras foram colocados nos respectivos poços e a corrida foi iniciada com uma voltagem de 130v durante 60 minutos. Após a corrida o gel foi observado e fotografado no transluminador UV *BioDoc-it* (UVP). Na figura 9 estão ilustradas as possíveis conjugações de fragmentos obtidos visualizados em gel.

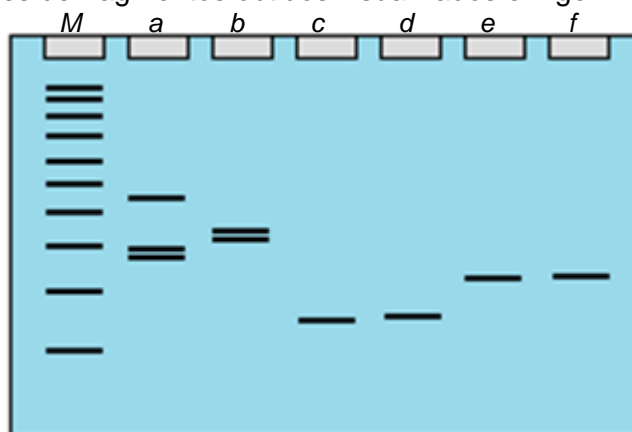


Figura 9: Representação de um gel de agarose com ilustração de fragmentos amplificados de diferentes tamanhos. Na coluna da letra *M* está representado o marcador molecular, com as bandas separadas por 100 pb, com a banda inferior de 100 pb de tamanho e a banda superior de 1000 pb de tamanho. Na coluna da letra *a* está representado o conjunto de possíveis fragmentos obtidos pela amplificação do ADN com o *multiplex I*, com o fragmento inferior de 282 ou 302 pb, o fragmento central de 299 pb e o fragmento superior de 444 pb. Na coluna da letra *b* está representado o conjunto de possíveis fragmentos obtidos pela amplificação do ADN com o

multiplex II, com o fragmento inferior de 312 pb e o fragmento superior de 351 pb. Na coluna da letra *c* está representado o fragmento de 155 pb obtido pela amplificação do ADN em *singleplex* com os *primers* F16094/R16248. Na coluna da letra *d* está representado o fragmento de 164 pb obtido pela amplificação do ADN em *singleplex* com os *primers* F15989/R16152. Na coluna da letra *e* está representado o fragmento de 217 pb obtido pela amplificação do ADN em *singleplex* com os *primers* F16533/RR180. Na coluna da letra *f* está representado o fragmento de 218 pb obtido pela amplificação do ADN em *singleplex* com os *primers* F402/R16152.

6. Purificação do ADN

O ADN resultante da amplificação por PCR que se pretendia sequenciar, foi purificado enzimaticamente através da utilização do método *ExoSAP-IT™* (USB), que consiste na remoção de *primers* e dNTPs não incorporados durante a amplificação do ADN. Por cada 5 µL de produto da PCR adicionaram-se 2 µL de *ExoSAP-IT™*, seguidamente procedeu-se a uma primeira incubação a 37°C durante 15 minutos no termociclador *iCycler* (BioRad) para activação enzimática, seguida pela sua inactivação através de uma segunda incubação a 80°C durante 15 minutos.

7. Sequenciação automática

O produto da PCR, purificado ou não purificado, que se pretendia sequenciar foi enviado para uma de duas empresas que prestam serviços de sequenciação do ADN, a Stabvida (Caparica, Portugal) ou a Macrogen (Seul, Coreia).

Para a empresa Stabvida foram enviados em separado 5 µL de produto da PCR e 5 µL de *primer* a 10 pmol/µL por reacção. Para a reacção de sequenciação, segundo as informações fornecidas pelo responsável, foram utilizados 20 ng de produto da PCR, *primer* a 3,3 pmol/µL e 8 µL de *BigDye® Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems), num volume total de reacção de 20 µL. A região analisada foi sequenciada apenas no sentido *forward*. A reacção de sequenciação ocorreu a uma temperatura de 96°C durante 1 minuto para desnaturação inicial do ADN, seguindo-se 25 ciclos de desnaturação a 96°C durante 10 segundos, hibridização a 55°C durante 6 segundos e extensão a 60°C durante 4 minutos. A remoção do corante *BigDye®* foi realizada através da utilização de placas Performa® DTR (EdgeBio) que contêm uma matriz de filtração em gel hidratada. O sequenciador utilizado foi o modelo ABI3730XL (Applied Biosystems).

Para a empresa Macrogen foram enviados em separado 5 µL de produto da PCR e 5 µL de *primer forward* a 10 pmol/µL por reacção. Na empresa Macrogen,

segundo as informações disponíveis no respectivo *website* (<http://www.macrogen.com>, 2011.10.01) a reacção foi realizada com condições *BigDye™ terminator cycling*, os produtos de reacção foram purificados com etanol e sequenciados no sequenciador ABI3730XL (Applied Biosystems).

8. Análise informática de sequências do ADN

As sequências do ADNmt obtidas por sequenciação automática dos diversos fragmentos do ADN foram analisadas e seleccionadas de acordo com os respectivos electroforetogramas, considerando a ausência/presença de sobreposição de picos, a definição dos picos, altura dos picos e presença/ausência de levantamento da linha de base, de modo a determinar-se a sequência do ADNmt com um nível de confiança elevado. As sequências representadas por electroforetogramas considerados de boa qualidade foram analisadas com o *software* BioEdit (Hall, 1999), procedendo-se à montagem das sequências dos fragmentos do ADN obtidas por sequenciação e construindo-se uma sequência consenso para cada amostra analisada, alinhando-se as sequências com a CRSr (Anderson *et al.*, 1981; Andrews *et al.*, 1999), determinando-se as diferenças entre as sequências (mutações) e o haplótipo do indivíduo dador de cada amostra.

9. Afiliação em haplogrupos

Os haplótipos determinados por sequenciação do ADN foram utilizados para a determinação do haplogrupo associado a cada amostra através de dois métodos diferentes apropriados ao estudo do ADNmt. Um dos métodos consistiu na utilização da ferramenta *mtDNA manager* (Lee *et al.*, 2008) actualizada a 2 de Outubro de 2011, que associa automaticamente as mutações a um determinado haplogrupo. O outro método foi a utilização da árvore filogenética *PhyloTree* (van Oven & Kayser, 2009) actualizada a 20 de Julho de 2011, determinando-se o haplogrupo de forma manual, através do percurso da árvore filogenética de acordo com uma sequência de acumulação de mutações.

III – Resultados

(Página intencionalmente deixada em branco)

1. Quantificação das amostras

1.1. Quantificação do ADN com fluorímetro

O ADN de algumas amostras de Goa seleccionadas aleatoriamente, foi quantificado através da utilização do fluorímetro *Qubit®* (Invitrogen). Foram quantificadas amostras cujo processo de extracção foi realizado há mais de um ano e amostras com o processo de extracção realizado recentemente (intervalo inferior a 6 meses). As amostras extraídas foram quantificadas obtendo-se valores de concentração muito semelhantes entre as amostras de Goa, tabela 4, com valor mínimo de 0,318 ng/μL valor máximo de 1,000 ng/μL e valor médio de 0,602 ng/μL. Verificou-se a presença do ADN nas amostras susceptível de ser amplificado por PCR.

Tabela 4 – Resultados da quantificação do ADN utilizando o fluorímetro. Na tabela encontra-se discriminado o período decorrido entre o processo de extracção e a análise das amostras extraídas. A fundo branco está a quantificação das amostras com um período de intervalo entre a extracção do ADN amplificação do ADN superior a 1 ano. A fundo cinzento estão as amostras com um período de intervalo entre a extracção do ADN amplificação do ADN recentemente (inferior a 6 meses).

Número da Amostra	Concentração [ng/μL]	Número da Amostra	Concentração [ng/μL]
Controlo positivo	0,323	108	0,653
3	0,612	108	0,378
7	0,318	109	0,852
8	0,401	110	0,588
16	0,717	110	0,375
17	0,748	111	0,515
21	0,792	112	1,000
26	0,617	113	0,478
30	0,738	113	0,462
30	0,518	114	0,444
88	0,759	115	0,560
88	0,506	116	0,539
89	0,877	117	0,815
90	0,728	124	0,391
90	0,692	123	0,394
91	0,632	125	0,710
94	0,709	133	0,690
100	0,526	134	0,462
107	0,530		

1.2. Quantificação por rt-qPCR

Para a realização da quantificação do ADN por PCR em tempo real testaram-se dois *kits* comerciais. No caso do *kit* comercial com o fluorocromo *SybrGreen* (página 42) verificou-se a presença de inibição quando a amostra foi utilizada sem diluição, não tendo ocorrido amplificação do ADN, sendo que quando a amostra é utilizada diluída (10x ou 100x) ocorre amplificação do ADN alvo (figura 10).

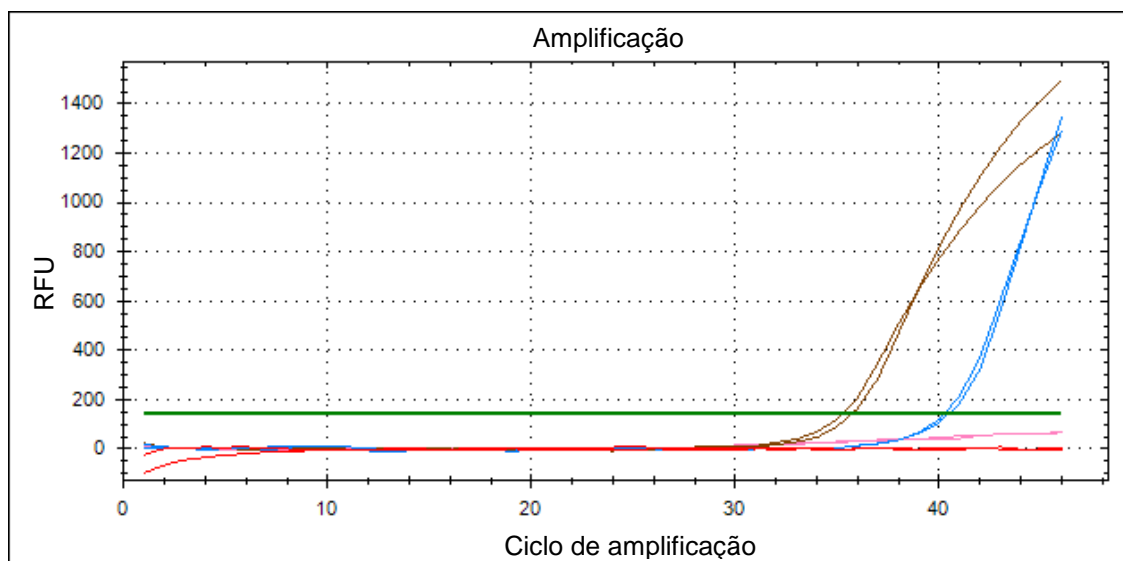


Figura 10: Resultados da amplificação por PCR em tempo real de uma amostra de Goa com diferentes diluições, com a utilização do *kit* comercial com o fluorocromo *SybrGreen*. No eixo das abcissas está representado o ciclo de amplificação e no eixo das ordenadas estão representadas as Unidades Relativas de Fluorescência (RFU). A vermelho estão representadas as duas réplicas do controlo negativo, a rosa estão representadas as duas réplicas da amostras de Goa não diluída, a castanho estão representadas as duas réplicas da amostra de Goa diluída 10x e a azul estão representadas as suas réplicas da amostra de Goa diluída 100x.

Com a utilização do *kit* comercial com o fluorocromo *EvaGreen* (página 42) não se verificou a presença de inibição, sendo visível que o número de ciclo de amplificação (amostra não diluída) < número de ciclo de amplificação (amostra diluída 10x) < número de ciclo de amplificação (amostra diluída 100x) (figura 11).

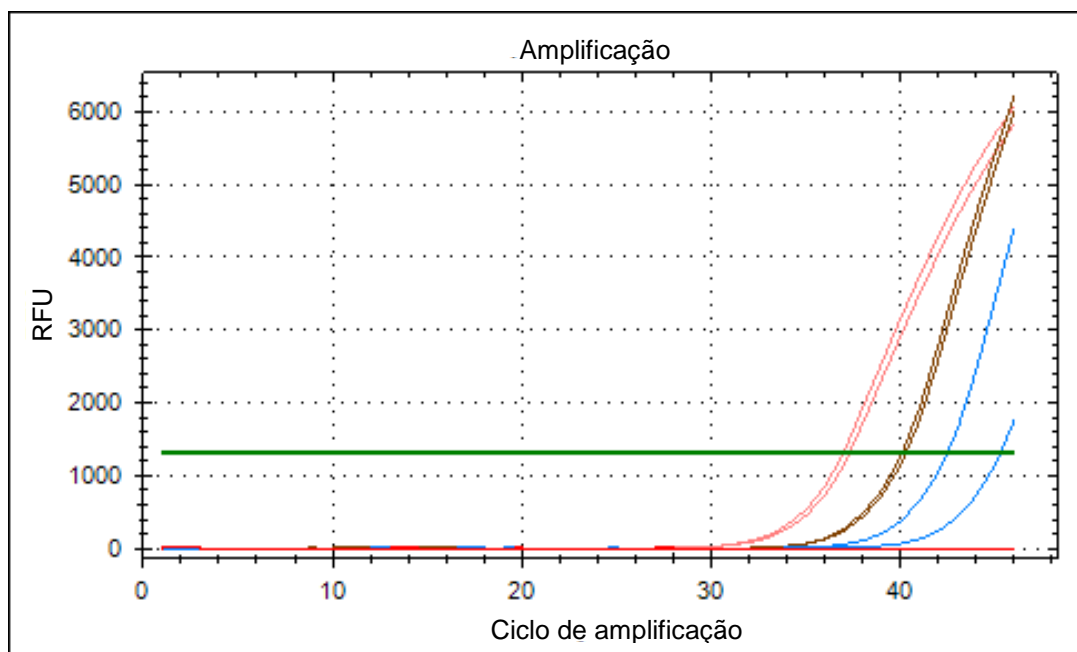


Figura 11: Resultados da amplificação por PCR em tempo real de uma amostra de Goa com diferentes diluições, com a utilização do *kit* comercial com o fluorocromo *EvaGreen*. No eixo das abcissas está representado o ciclo de amplificação e no eixo das ordenadas estão representadas as Unidades Relativas de Fluorescência (RFU). A vermelho estão representadas as duas réplicas do controlo negativo, a rosa estão representadas as duas réplicas da amostras de Goa não diluída, a castanho estão representadas as duas réplicas da amostra de Goa diluída 10x e a azul estão representadas as suas réplicas da amostra de Goa diluída 100x.

Com o fluorocromo *SybrGreen* procedeu-se também à amplificação do ADN por PCR em tempo real de uma série de diluições decimais da amostra de Goa número 30 com 2 intervalos diferentes decorridos entre o processo de extracção e amplificação. No caso do intervalo superior a um ano não ocorreu amplificação do ADN em 3 réplicas da amostra não diluída e em 3 réplicas da amostra diluída 10x obteve-se Cq médio de 36,25. No caso do intervalo inferior a 6 meses o valor do ciclo de amplificação em 2 réplicas da amostra foi de 34,61 e em 3 réplicas da amostra diluída 10x obteve-se o valor do ciclo de amplificação médio de 34,19.

Para reduzir a influência de inibidores utilizou-se o *kit* comercial com o fluorocromo *EvaGreen* para a quantificação das amostras. A curva padrão construída para quantificação das amostras é caracterizada por uma eficiência de 108,4%, um declive de -3,1368, uma intercepção com o eixo ordenadas de 44,439 e um r^2 de 0,999 (figura 12).

Foi realizada a quantificação de um fragmento de 164 pb, inserido na região HRVI do ADNmt, produzido pelos *primers* F15989/R16152, em 8 amostras: 1 amostra controlo positivo (C+) e 7 amostras de Goa (3, 30, 107, 109, 110, 115, 123). Foi também realizada a reacção com 2 **controles negativos** que não apresentaram sinal de amplificação.

As amostras 107, 109 e 123 revelaram Cq e Log(Quantidade Inicial) que não se enquadram na recta padrão, o que reflecte uma quantidade bastante reduzida, não tendo sido por isso calculado o número de cópias inicial. As restantes amostras (C+, 3, 30, 110 e 115) enquadram-se na recta padrão definida, tendo sido por isso determinado o número de cópias/quantidade inicial.

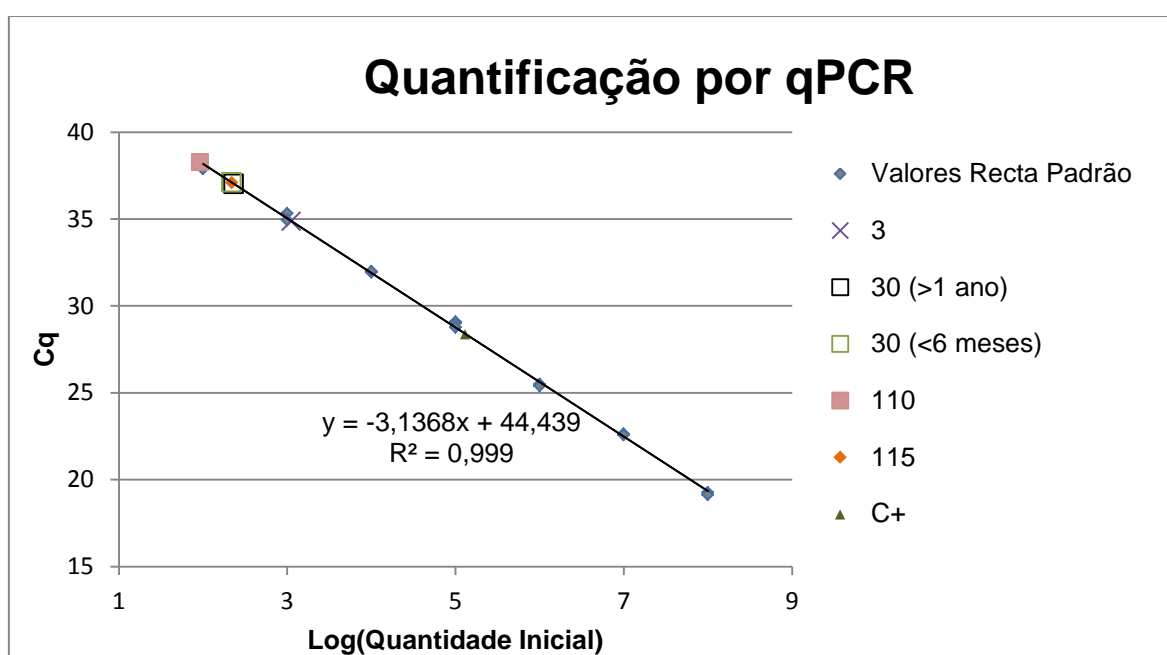


Figura 12: Resultados da quantificação do fragmento de 164 pb inserido na região HVRI do ADNmt. No eixo das abcissas está representado o logaritmo do número de cópias iniciais (Quantidade Inicial) do ADN e no eixo das ordenadas está representado o ciclo de quantificação (Cq). Na amostra 30, entre parêntesis, está especificado o intervalo entre os processos de extracção/amplificação do ADN, as restantes amostras de Goa possuem um intervalo superior a um ano.

Como representado na tabela 5, a amostra **controlo positivo** apresenta o maior número de cópias (130.526 cópias), seguida pela amostra 3 (1.123 cópias), com as amostras 30 e 115 com valores na ordem das 200 cópias e a amostra 123 apresenta o menor número de cópias (44 cópias). A amostra 30 foi amplificada duas vezes, uma onde se utilizou ADN extraído a mais de um ano e outra onde se utilizou ADN extraído num

intervalo inferior a 6 meses, verificando-se resultados semelhantes, 232 cópias e 219 cópias, respectivamente.

Tabela 5 – Resultados da quantificação do fragmento de 164 pb inserido na região HVRI do ADNmt. O número de cópias inicial foi calculado através da equação da recta padrão. “Nd” indica o número de cópias não determinado. Na amostra 30, entre parêntesis, está especificado o intervalo entre os processos de extracção/amplificação do ADN, as restantes amostras de Goa possuem um intervalo superior a um ano.

Amostra	y (Cq)	x (log(QI))	Número de cópias inicial
Controlo positivo	28,39	5,115696	130.526
3	34,87	3,050561	1123
30 (>1 ano)	37,02	2,365149	232
30 (<6 meses)	37,1	2,339645	219
107	40,36	1,30037	Nd
109	43,67	0,245154	Nd
110	38,27	1,966654	93
115	37,09	2,342833	220
123	39,27	1,647858	Nd

Para além das amostras de Goa aqui referidas foram quantificadas por PCR em tempo real mais 17 amostras de Goa, contudo as duas réplicas do **controlo negativo** demonstraram amplificação do ADN após o ciclo de amplificação número 40, bastante próximo dos ciclos de amplificação das amostras analisadas, Cq entre 35,57 e 39,11, não se tendo por este motivo considerado estas amostras para análise.

2. Amplificação e sequenciação da região não codificante

2.1. Amplificação de todas as amostras de Goa

O ADN das amostras da colecção de Goa com o processo de extracção realizado há mais de um ano foi amplificado por PCR com o *multiplex* I e *multiplex* II, de forma a avaliar-se o estado de todas as amostras já extraídas e perceber qual a possibilidade de obtenção de uma sequência completa em cada caso.

2.1.1. Utilização do *multiplex* II

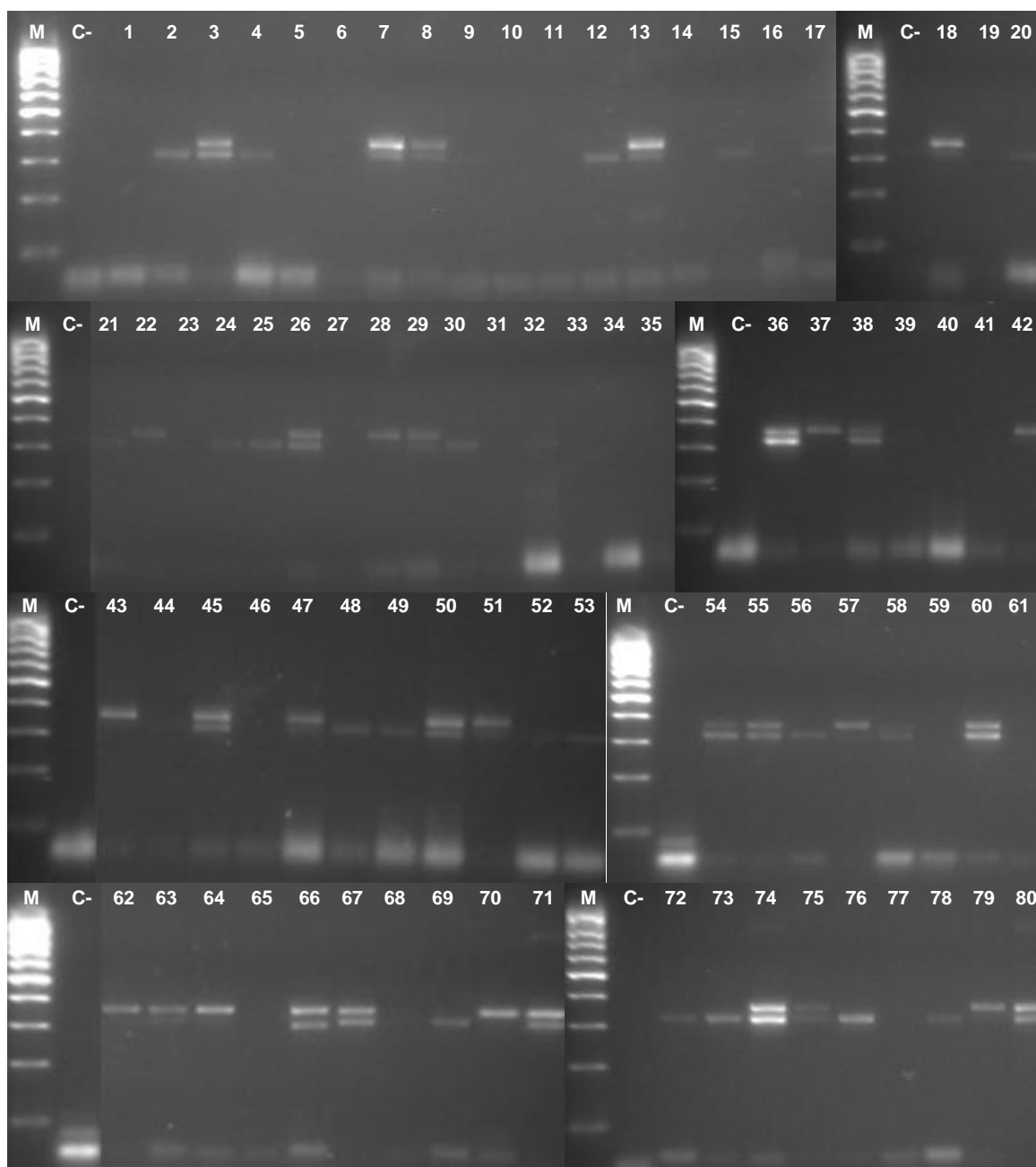
O primeiro *multiplex* utilizado foi o *multiplex* II (página 45) devido à maior possibilidade de sucesso de amplificação de ambos os fragmentos alvo e à maior facilidade de visualização das bandas. Foram analisadas 127 amostras (amostras nº: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126 e 127) (figura 13).

De todas as amostras amplificadas por PCR, em 30 não ocorreu amplificação do ADN (amostras nº: 1, 5, 6, 9, 10, 11, 14, 16, 19, 23, 27, 31, 33, 34, 35, 39, 40, 41, 46, 59, 61, 65, 68, 77, 82, 83, 85, 106, 120 e 126); em 37 amostras está presente apenas a banda de 312 pb, correspondente ao fragmento do ADN amplificado com os *primers* F16197/R16509 (amostras nº: 2, 4, 12, 15, 17, 20, 21, 24, 25, 30, 32, 44, 48, 49, 52, 53, 56, 69, 72, 73, 76, 78, 81, 84, 88, 89, 91, 93, 95, 97, 98, 99, 112, 118, 119, 122 e 125); em 15 amostras está presente apenas a banda de 351 pb correspondente ao fragmento do ADN amplificado com os *primers* F109/R460 (amostras nº: 22, 28, 37, 42, 43, 57, 62, 63, 64, 70, 79, 92, 102, 108 e 114); em 45 amostras ambas as bandas estão presentes, nomeadamente a banda correspondente ao fragmento do ADN de 312 pb amplificado com os *primers* F16197/R16509 e a banda correspondente ao fragmento do ADN de 351 pb amplificado com os *primers* F109/R460 (amostras nº: 3, 7, 8, 13, 18, 26, 29, 36, 38, 45, 47, 50, 51, 54, 55, 58, 60, 66, 67, 71, 74, 75, 80, 86, 87, 90, 94, 96, 100, 101, 103, 104, 105, 107, 109, 110, 111, 113, 115, 116, 117, 121, 123, 124 e 127).

Em 4 amostras (amostras nº: 71, 74, 80 e 117) está presente no gel, uma banda de aproximadamente 850 pb, que não corresponde a nenhum dos fragmentos alvo.

Juntamente com as amostras de Goa também foi amplificada uma amostra **controlo positivo** (C+), de colheita recente de saliva em cartão FTA e extracção recente, na qual se verificou a amplificação dos 2 fragmentos alvo, com a visualização de 2 bandas, uma de 312 pb resultante da amplificação com os *primers* F16197/R16509 e outra de 352 pb resultante da amplificação com os *primers* F109/R460.

Em todas as reacções de amplificação realizadas o **controlo negativo** (C-) não indica amplificação do ADN, não estando presentes bandas no gel.



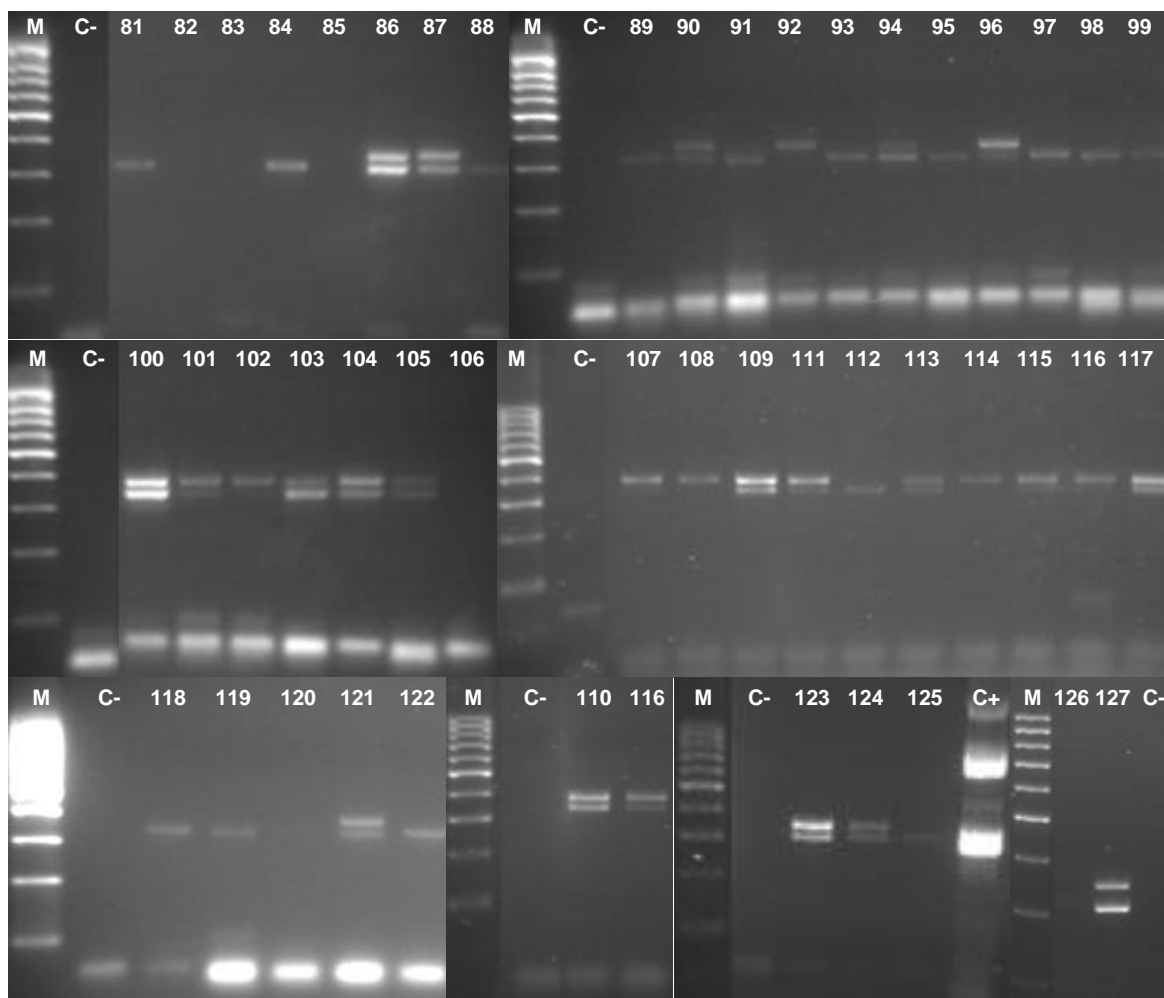


Figura 13: Fotografias de 16 géis de agarose 5% de reacções com o *multiplex* II. A letra M representa o marcador molecular de 100 pb; C- (controlo negativo) representa a reacção onde a amostra foi substituída por água *Mili-Q* autoclavada; C+ (controlo positivo) representa a reacção onde foi utilizada uma amostra que não corresponde à colecção de Goa; e os algarismos representam o número atribuído às amostras de Goa. As reacções de todas as amostras ocorreram com os *primers* que constituem o *multiplex* II, nomeadamente, F16197/R16509 (312 pb) e F109/R460 (351 pb); as reacções ocorreram num volume total de 25 µL, com 1 µL do ADN/H₂O.

2.1.2. Utilização do *multiplex* I

As amostras que apresentam ambas as bandas com o *multiplex* II no gel de agarose, nomeadamente a banda de 312 pb resultante da amplificação com os *primers* F16197/R16509 e a banda de 351 pb resultante da amplificação com os *primers* F109/R460 foram analisadas com o *multiplex* I (página 45), não se tendo analisado as amostras 51 e 75 devido à reduzida intensidade das bandas. As amostras 108 e 125 foram também analisadas com o *multiplex* I, apesar de apenas ter ocorrido amplificação

do ADN com os *primers* F109/R460 (351 pb) na amostra 108, e com os *primers* F16197/R16509 (312 pb) na amostra 125, nomeadamente 45 amostras (amostras nº: 3, 7, 8, 13, 18, 26, 29, 36, 38, 45, 47, 50, 54, 55, 60, 66, 67, 71, 74, 80, 86, 87, 90, 94, 96, 100, 101, 103, 104, 105, 107, 108, 109, 110, 111, 112, 113a/b, 115, 116, 117, 121, 123, 124, 125 e 127) (figura 14).

Das amostras analisadas, em 5 não ocorreu amplificação do ADN, não se visualizando portanto qualquer banda no gel de agarose (amostras nº: 45, 74, 86, 96 e 108); em 27 amostras (amostras nº: 18, 29, 36, 38, 47, 50, 54, 55, 60, 66, 67, 71, 80, 87, 90, 104, 107, 110, 111, 113a/b, 115, 116, 117, 121, 124, 125 e 127) estão presentes apenas uma ou duas das bandas que representam os 3 fragmentos alvo, um fragmento de 444 pb correspondente à amplificação do ADN com os *primers* F15989/R16433, um fragmento de 299 pb correspondente à amplificação do ADN com os *primers* F16450/R180 e um fragmento de 282 pb correspondente à amplificação do ADN com os *primers* F317/R599; em 13 amostras (amostras nº: 3, 7, 8, 13, 26, 94, 100, 101, 103, 105, 109, 112 e 123) estão presentes os 3 fragmentos alvo, um fragmento de 299 pb correspondente à amplificação do ADN com os *primers* F16450/R180 e um fragmento de 282 pb correspondente à amplificação do ADN com os *primers* F317/R599.

Em 9 amostras (amostras nº: 3, 8, 13, 18, 36, 100, 109, 112 e 123) está presente no gel, uma banda de aproximadamente 750 pb, que não corresponde a nenhum dos fragmentos alvo.

Juntamente com as amostras de Goa também foi amplificada uma amostra **controlo positivo** (C+), de colheita recente de saliva em cartão FTA e extracção recente, na qual se visualiza 3 bandas correspondentes à amplificação dos 3 fragmentos alvo, um fragmento de 299 pb correspondente à amplificação do ADN com os *primers* F16450/R180 e um fragmento de 282 pb correspondente à amplificação do ADN com os *primers* F317/R599, também é possível visualizar uma quarta banda, de aproximadamente 700 pb, que não corresponde a nenhum dos fragmentos alvo.

Em todas as reacções de amplificação realizadas o **controlo negativo** (C-) não indica amplificação do ADN, não estando presentes bandas no gel.

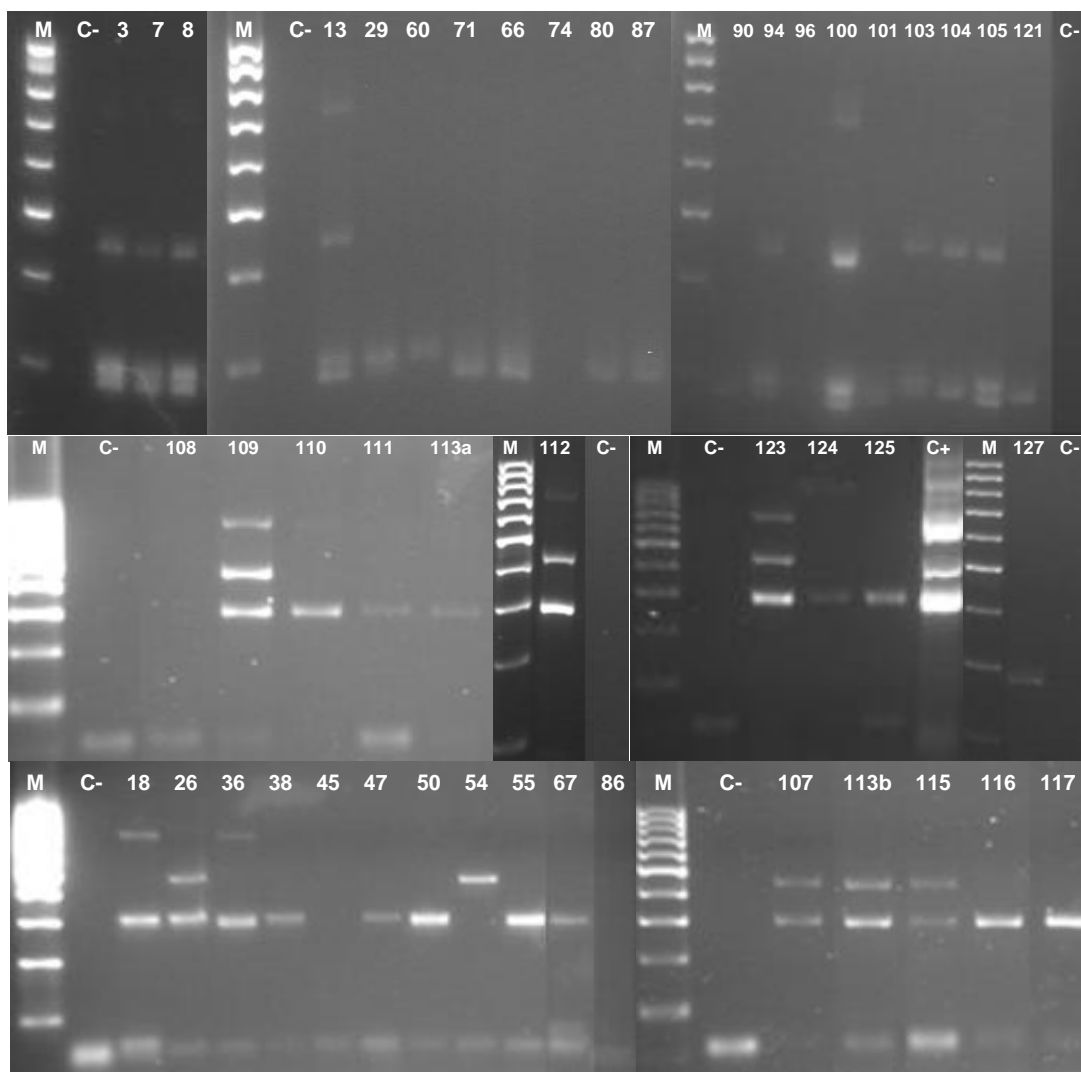


Figura 14: Fotografias de 7 géis de agarose 5% de reacções com o *multiplex I*. A letra M representa o marcador molecular de 100 pb; C- (controlo negativo) representa a reacção onde a amostra foi substituída por água *Mili-Q* autoclavada; C+ (controlo positivo) representa a reacção onde foi utilizada uma amostra que não corresponde à colecção de Goa; e os algarismos representam o número atribuído às amostras de Goa. As reacções de todas as amostras ocorreram com os *primers* que constituem o *multiplex I*, nomeadamente, F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb); as reacções ocorreram num volume total de 25 μ L, com 1 μ L do ADN/H₂O.

Combinando os dados obtidos pela amplificação do ADN por PCR e visualização dos produtos da PCR em gel de agarose (tabela 6), é perceptível que apenas em uma pequena percentagem de amostras (10%) seria hipoteticamente possível obter a sequência completa da região não codificante do ADNmt, o que equivale a 12 amostras em 127. No caso do *multiplex I* a percentagem de amostras sem amplificação do ADN é de 11% ocorrendo amplificação dos 3 fragmentos alvo apenas em 29% das

amostras. Já no *multiplex* II a percentagem de amostras sem amplificação é superior (24%), tal como a percentagem de amostras onde ambos os fragmentos alvo foram amplificados (35%). Assim, apesar da percentagem de amostras sem amplificação do ADN ser inferior a 25%, a percentagem mais elevada é a de obtenção incompleta dos fragmentos que formam cada *multiplex*, de 60% no *multiplex* I e de 41% no *multiplex* II, o que impossibilita a obtenção de toda a região não codificante do ADNmt.

Tabela 6 – Síntese dos resultados das amplificações das amostras de Goa com o *multiplex* I e *multiplex* II. Nas colunas designadas por *Multiplex* I e *Multiplex* II está indicado o número de amostras e a sua percentagem associadas à característica indicada na primeira coluna da tabela.

	<i>Multiplex</i> I	<i>Multiplex</i> II
Apenas F16197/R16509 (312 pb) ou F109/R460 (351 pb)	-	52 (41%)
F16197/R16509 (312 pb) e F109/R460 (351 pb)	-	45 (35%)
Apenas 1 ou 2 de F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb)	27 (60%)	-
F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb)	13 (29%)	-
Amostras sem amplificação	5 (11%)	30(24%)
Número total de amostras analisadas	45 (100%)	127 (100%)
Amostras com 5 fragmentos alvo amplificados	13 (10%)	

2.2. Amplificação de amostras com extracção recente

2.2.1. Utilização do *multiplex* II

Foram realizadas novas extracções de 28 amostras da colecção de Goa (amostras nº: 9, 20, 24a/b, 30, 32, 35, 36, 41, 53, 55, 62, 66, 86, 87, 96, 108, 110, 112, 113, 125, 128, 129, 131, 132, 133, 134, 135 e 136), realizando-se a sua extracção, primeiramente, com o *multiplex* II (figura 15). Em 2 amostras (amostras nº: 35 e 87) não ocorreu amplificação do ADN. Em 7 amostras (amostras nº: 9, 41, 62, 66, 86, 108 e 125) está apenas presente a banda de 312 pb correspondente ao fragmento do ADN amplificado com os *primers* F16197/R16509. Em 3 amostras (amostras nº: 55, 96 e 112) apenas está presente a banda de 351 pb correspondente ao fragmento do ADN amplificado com os *primers* F109/R460. Em 16 amostras (amostras nº: 20, 24a/b, 30, 32, 36, 53, 110, 113, 128, 129, 131, 132, 133, 134, 135 e 136) ambas as bandas estão presentes nomeadamente a banda correspondente ao fragmento do ADN de 312 pb

amplificado com os *primers* F16197/R16509 e a banda correspondente ao fragmento do ADN de 351 pb amplificado com os *primers* F109/R460. Em 3 amostras (amostras nº: 20, 24a e 32) está presente no gel uma banda de aproximadamente 850 pb, que não corresponde a nenhum dos fragmentos alvo.

Em todas as reacções de amplificação realizadas, o controlo negativo (C-) não indica amplificação do ADN, não estando presentes bandas no gel.

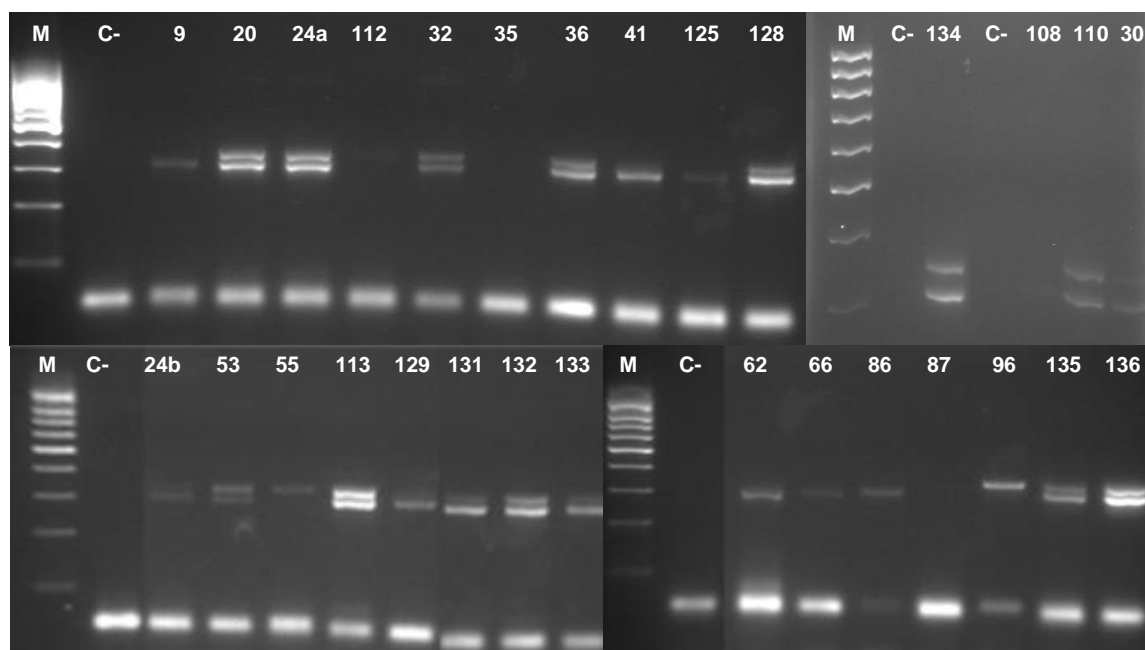


Figura 15: Fotografias de 4 géis de agarose 5% de reacções com *multiplex II*. A letra M representa o marcador molecular de 100pb; C- (controlo negativo) representa a reacção onde a amostra foi substituída por água *Mili-Q* autoclavada; e os algarismos representam o número atribuído às amostras na colecção de Goa. As reacções de todas as amostras ocorreram com os *primers* que constituem o *multiplex II*, nomeadamente, F16197/R16509 (312 pb) e F109/R460 (351 pb); as reacções ocorreram num volume total de 25 µL, com 1 µL do ADN/H₂O.

2.2.2. Utilização do *multiplex I*

As amostras que apresentam ambas as bandas com o *multiplex II* no gel de agarose, nomeadamente a banda de 312 pb resultante da amplificação com os *primers* F16197/R16509 e a banda de 351 pb resultante da amplificação com os *primers* F109/R460 foram analisadas com o *multiplex I*, nomeadamente 16 amostras (amostras nº: 20, 24, 30, 32, 36, 108, 110, 113, 128, 129, 131, 132, 133a/b, 134, 135 e 136) (figura 16).

Das amostras analisadas ocorreu em todas a amplificação de pelo menos um fragmento alvo; em 7 amostras (amostras nº: 20, 24, 32, 36, 128, 129 e 131) estão

presentes apenas uma ou duas das bandas que representam os 3 fragmentos alvo, um fragmento de 444 pb correspondente à amplificação do ADN com os *primers* F15989/R16433, um fragmento de 299 pb correspondente à amplificação do ADN com os *primers* F16450/R180 e um fragmento de 282 pb correspondente à amplificação do ADN com os *primers* F317/R599; em 9 amostras (amostras nº: 30, 108, 110, 113, 132, 133a/b, 134, 135 e 136) estão presentes os 3 fragmentos alvo um fragmento de 299 pb correspondente à amplificação do ADN com os *primers* F16450/R180 e um fragmento de 282 pb correspondente à amplificação do ADN com os *primers* F317/R599. Na amostra 30 está presente no gel uma banda de aproximadamente 750 pb, que não corresponde a nenhum dos fragmentos alvo. Na amostra 133, amplificada duas vezes (a/b), verificou-se a existência de diferença de intensidade.

Em todas as reacções de amplificação realizadas, o controlo negativo (C-) não indica amplificação do ADN, não estando presentes bandas no gel.

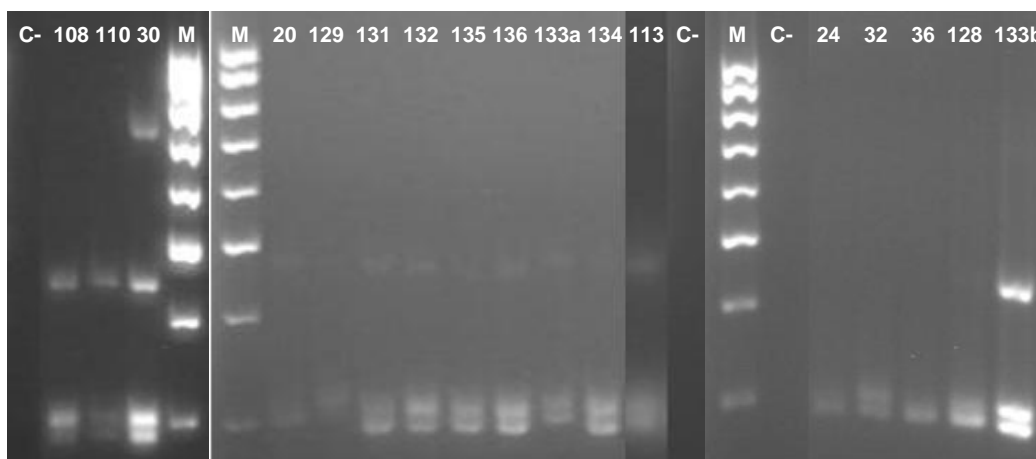


Figura 16: Fotografias de 3 géis de agarose a 5% de reacções com *multiplex* I. A letra M representa o marcador molecular de 100 pb; C- (controlo negativo) representa a reacção onde a amostra foi substituída por água *Mili-Q* autoclavada; e os algarismos representam o número atribuído às amostras da colecção de Goa. As reacções de todas as amostras ocorreram com os *primers* que constituem o *multiplex* I, nomeadamente, F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb); as reacções ocorreram num volume total de 25 µL, com 1 µL do ADN/H₂O.

Combinando os dados obtidos pela amplificação do ADN por PCR e visualização dos produtos da PCR em gel de agarose (tabela 7), é perceptível que apenas em 9 de 28 amostras analisadas (32%) seria hipoteticamente possível obter a sequência completa da região não codificante do ADNmt. No caso do *multiplex* I a percentagem de amostras sem amplificação do ADN é de 0% ocorrendo amplificação dos 3 fragmentos alvo em 56% das amostras, percentagem superior à amplificação de apenas 1 ou 2 dos 3

fragmentos alvo (44%). Já no *multiplex* II a percentagem de amostras sem amplificação é de 7%, com a percentagem de amostras onde ambos os fragmentos alvo foram amplificados de 57%, superior à percentagem de amostras onde apenas um fragmento alvo foi amplificado (36%).

É perceptível que o número de amostras onde não ocorreu amplificação de qualquer fragmento é reduzido, sendo a maior percentagem representativa de amostras com amplificação de todos os fragmentos alvo (56% ou 57%).

Tabela 7 – Síntese dos resultados das amplificações das amostras de Goa com o *multiplex* I e *multiplex* II. Nas colunas designadas por *Multiplex* I e *Multiplex* II está indicado o número de amostras e a percentagem equivalente.

	<i>Multiplex</i> I	<i>Multiplex</i> II
Apenas F16197/R16509 (312 pb) ou F109/R460 (351 pb)	-	10 (36%)
F16197/R16509 (312 pb) e F109/R460 (351 pb)	-	16 (57%)
Apenas 1 ou 2 de F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb)	7 (44%)	-
F15989/R16433 (444 pb), F16450/R180 (299 pb) e F317/R599 (282 pb)	9 (56%)	-
Amostras sem amplificação	0 (0%)	2 (7%)
Número total de amostras analisadas	16 (100%)	28 (100%)
Amostras com 5 fragmentos alvo amplificados	9	(32%)

2.2.3. Comparação amostras com intervalo extracção/amplificação superior a 1 ano e amostras com intervalo extracção/amplificação inferior a 6 meses

De forma geral, quando se comparam as percentagens obtidas em diferentes intervalos de extracção do ADN verifica-se que melhores resultados são obtidos com a extracção recente (tabela 8).

Relativamente ao *multiplex* I, a extracção do ADN realizada recentemente é caracterizada por maior percentagem (56%) de amostras com amplificação bem-sucedida de todos os fragmentos alvo, comparativamente à extracção do ADN superior a 1 ano (29%) e por outro lado menor percentagem (0%) de amostras onde não ocorreu amplificação de nenhum dos fragmentos alvo no caso da extracção recente, contra 11% na extracção num intervalo superior a 1 ano.

Relativamente ao *multiplex* II, a extracção do ADN realizada recentemente é caracterizada por maior percentagem (57%) de amostras com amplificação bem-sucedida de todos os fragmentos alvo, comparativamente à extracção do ADN superior a 1 ano

(35%) e por outro lado menor percentagem (7%) de amostras onde não ocorreu amplificação de nenhum dos fragmentos alvo no caso da extracção recente, contra 24% na extracção num intervalo superior a 1 ano.

Tabela 8 – Comparação geral dos resultados da amplificação das amostras de Goa com o *multiplex* I e *multiplex* II com diferentes intervalos decorridos entre o processo de extracção e a análise das amostras extraídas.

	<i>Multiplex</i> I		<i>Multiplex</i> II	
	Intervalo superior a 1 ano	Intervalo inferior a 6 meses	Intervalo superior a 1 ano	Intervalo inferior a 6 meses
Amplificação incompleta dos fragmentos alvo	27 (60%)	7 (44%)	52 (41%)	10 (36%)
Amplificação de todos os fragmentos alvo	13 (29%)	9 (56%)	45 (35%)	16 (57%)
Amostras sem amplificação	5 (11%)	0 (0%)	30(24%)	2 (7%)

Quando se procede à comparação individual das amostras extraídas em diferentes intervalos temporais, é visível a variabilidade de resultados (tabela 9). Em 9 amostras obteve-se o mesmo resultado com os diferentes intervalos, nomeadamente nas amostras 35, 36, 62, 108, 110, 112, 113 e 125 do *multiplex* II e na amostra 36 do *multiplex* I. No caso do *multiplex* II em 12 amostras onde se verificaram diferenças, 7 apresentaram melhores resultados com a extracção recente e 5 melhores resultados com a extracção superior a 1 ano. No *multiplex* I nas 3 amostras onde se verificaram diferenças, os melhores resultados foram obtidos com a extracção recente.

Tabela 9 – Comparação entre amostras com intervalo de extracção/amplificação superior a 1 ano e inferior a 6 meses. O algarismo "0" representa ausência de amplificação, o símbolo "-" representa amplificação incompleta dos fragmentos alvo, o símbolo "+" representa amplificação de todos os fragmentos alvo e "N.A." representa uma amostra não amplificada com o respectivo *multiplex* e intervalo.

Amostra	Multiplex I		Multiplex II	
	Intervalo superior a 1 ano	Intervalo inferior a 6 meses	Intervalo superior a 1 ano	Intervalo inferior a 6 meses
9	N.A.	N.A	0	-
20	N.A	-	-	+
24	N.A	-	-	+
30	N.A	+	-	+
32	N.A	-	-	+
35	N.A	N.A	0	0
36	-	-	+	+
41	N.A	N.A	0	-
53	N.A	N.A	-	+
55	N.A	N.A	+	-
62	N.A	N.A	-	-
66	N.A	N.A	+	-
86	N.A	N.A	+	-
87	N.A	N.A	+	0
96	N.A	N.A	+	-
108	0	+	-	-
110	-	+	+	+
112	N.A	N.A	-	-
113	-	+	+	+
125	N.A	N.A	-	-

2.3. Sequenciação de amostras teste

O produto da PCR de algumas amostras de Goa com processo de extracção e amplificação separados por 1 intervalo superior a 1 ano (amostras nº: 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 123, 124 e 125), foi enviado para sequenciação na empresa Stabvida, na tentativa de obtenção da sequência da região não codificante de cada amostra. Não foi realizada purificação previamente ao envio dos produtos da PCR enviados para sequenciação, mas devido à dificuldade na obtenção bons de resultados de sequenciação, alguns dos produtos da PCR foram purificados na empresa onde se

realizou a sequenciação e os produtos da PCR das amostras 123, 124 e 125 foram enviados para sequenciação após purificação pelo método enzimático ExoSAP-IT, tendo-se verificado a melhoria da qualidade dos electroforetogramas.

Da sequenciação do ADN obtiveram-se resultados considerados de boa qualidade, caracterizados por electroforetogramas com picos bem definidos e de alturas semelhantes entre si, sem/reduzida sobreposição de picos e sem/reduzido levantamento da linha de base (figura 17).

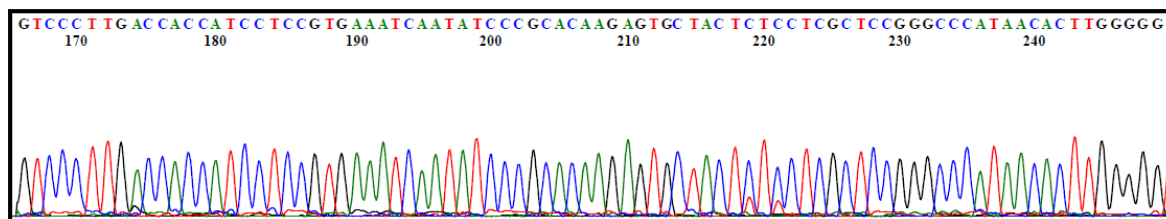


Figura 17: Exemplo de um electroforetograma de boa qualidade obtido por sequenciação de um fragmento do ADNmt de uma amostra de Goa.

Na sequenciação com o *primer* F15989 obtiveram-se sequências de qualidade com as amostras número 109, 112, 115 e 123, com sequências de 733 pb, 483 pb, 425 pb e 419 pb, respectivamente. Na sequenciação com o *primer* F16197 obtiveram-se sequências de qualidade com as amostras número 110, 112, 115, 117, 123, 124 e 125, com sequências de 284 pb, 284 pb, 297 pb, 293 pb, 290 pb, 270 pb e 288 pb, respectivamente. Na sequenciação com o *primer* F16450 obtiveram-se sequências de qualidade com as amostras número 107, 109, 110, 112, 114, 115, 117 e 124, com sequências de 274 pb, 273 pb, 280 pb, 278 pb, 272 pb, 277 pb, 280 pb e 271 pb, respectivamente. Na sequenciação com o *primer* F109 obtiveram-se sequências de qualidade com as amostras número 107, 108, 109, 113, 115, 116, 117, 123 e 124, com sequências de 330 pb, 330 pb, 330 pb, 337 pb, 338 pb, 335 pb, 333 pb, 330 pb e 330 pb, respectivamente. Na sequenciação com o *primer* F317 obtiveram-se sequências de qualidade com as amostras número 109, 110, 115, 117, 123, 124 e 125, com sequências de 275 pb, 274 pb, 288 pb, 260 pb, 254 pb, 260 pb e 255 pb. Com a amostra **controle positivo** obtiveram-se resultados de qualidade para os *primers* utilizados, a excepção do *primer* F109.

Para além de sequências caracterizadas por electroforetogramas de qualidade, da sequenciação do ADN também se obtiveram electroforetogramas de reduzida qualidade, principalmente devida à elevada sobreposição de picos (figura 18). Neste caso os resultados não são apresentados uma vez que não foram considerados para a determinação das sequências das amostras.

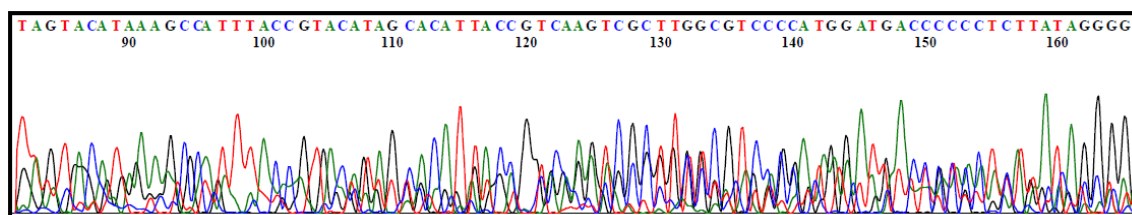


Figura 18: Exemplo de um electroforetograma de má qualidade obtido por sequenciação de um fragmento do ADNmt de uma amostra de Goa.

Na tabela 10 estão representadas as regiões das sequências das amostras de Goa que apresentaram melhor qualidade no respectivo electroforetograma, com a sua associação à respectiva região hipervariável do ADNmt, de acordo com a sequência obtida para cada região.

Na região HVRI não foi possível obter a totalidade da sequência em nenhuma das amostras, sendo que em 3 das amostras (amostras nº: 109, 115 e 123) apenas estão em falta 20 pb que correspondem à porção inicial da região HVRI, havendo mais 6 amostras (amostras nº: 107, 110, 112, 117, 124 e 125) com sequência incompleta e 4 amostras (amostras nº: 108, 113, 114 e 116) onde não se obteve qualquer resultado para a região HVRI.

Na região HVRII foi possível obter a totalidade da sequência em 4 amostras (amostras nº: 108, 109, 115 e 117), não se tendo obtido a qualquer resultado apenas com uma amostra (amostra nº: 125), com as restantes 8 amostras (amostras nº: 107, 110, 112, 113, 114, 116, 123 e 124) com sequências incompletas.

Na região HVRIII foi possível obter a totalidade da sequência em 5 amostras (amostras nº: 115, 117, 123, 124 e 125), em 4 amostras (amostras nº: 108, 109, 110 e 116) obteve-se sequência incompleta e em 4 amostras (amostras nº: 107, 112, 113 e 114) não se obteve qualquer resultado.

Tabela 10 – Síntese dos resultados de sequenciação das amostras de Goa. Estão representados os intervalos das regiões das sequências de cada amostra com resultados de melhor qualidade visíveis no respectivo electroforetograma, com a associação dos resultados às respectivas regiões hipervariáveis, HVRI (16.024 pb a 16.356 pb), HVRII (73 pb a 340 pb) e HVRIII (438 pb a 574 pb) e classificação das regiões de acordo com a informação obtida em ausente (x), incompleta (○) e completa (●). (*) representa aproximadamente 20 pb em falta na porção inicial da sequência do ADN.

Amostra	Regiões	HVRI	HVRII	HVRIII
107	16511-16569, 1-302	○	○	x
108	153-459	x	●	○
109	16039-16569, 1-459	○*	●	○
110	1-179, 372-491, 16237-16569	○	○	○
112	1-180, 16129-16569	○	○	x
113	153-299	x	○	x
114	1-180, 16504-16569	x	○	x
115	1-619, 16032-16509, 16525-16569	○*	●	●
116	151-460	x	○	○
117	8-619, 16239-16509	○	●	●
123	169-588, 16032-16498	○*	○	●
124	1-178, 183-588, 16255-16496, 16534-16569	○	○	●
125	371-588, 16255-16497	○	x	●

2.4. Amplificação e sequenciação de fragmentos curtos das amostras teste

2.4.1. Amplificação em singleplex das amostras

Para completar as sequências do ADN obtidas anteriormente por sequenciação, procedeu-se à amplificação em *singleplex* do ADN de algumas das amostras de Goa já analisadas (amostras nº:123, 109, 124, 115 e 117) e também da amostra C+ (controlo positivo).

Nas amostras analisadas agora em *singleplex* apenas a amostra 115 falhou com um dos *primers*, tendo todas as outras amostras tido uma amplificação bem-sucedida com todos os *primers* utilizados (figura 19). Na amostra controlo positivo verifica-se também ocorreu amplificação do fragmento alvo. O controlo negativo não apresenta qualquer banda.

Em todas as reacções de amplificação realizadas o **controlo negativo** (C-) não indica amplificação do ADN, não estando presentes bandas no gel de agarose.

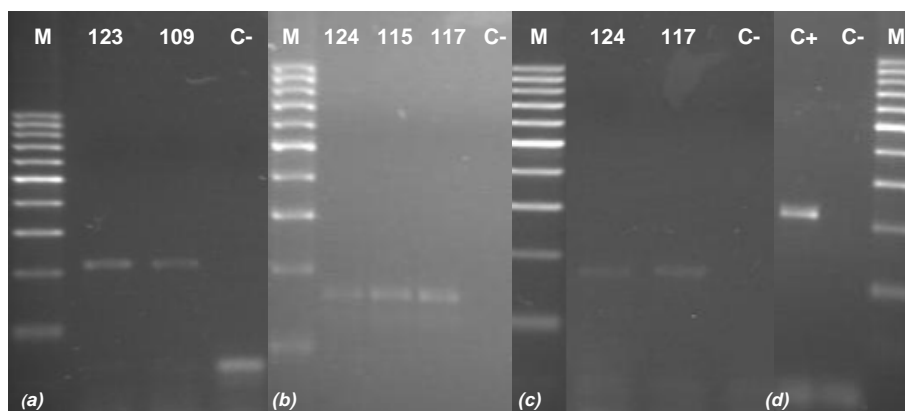


Figura 19: Fotografias de 4 géis de agarose 5% de reacções *singleplex*. A letra M representa o marcador molecular de 100pb; C- (controlo negativo) representa a reacção onde a amostra foi substituída por água *Mili-Q* autoclavada; e C+ (controlo positivo) representa a reacção onde foi utilizada uma amostra que não corresponde à colecção de Goa; os algarismos representam o número atribuído às amostras na colecção de Goa. **(a)** Reacção em *singleplex* com os *primers* F16533/R180 (217 pb); a reacção ocorreu num volume total de 25 µL, com 1 µL do ADN/H₂O. **(b)** Reacção em *singleplex* com os *primers* F15989/R16152 (164 pb); a reacção ocorreu num volume total de 25 µL, com 2 µL do ADN/H₂O. **(c)** Reacção em *singleplex* com os *primers* F16094/R16248 (155 pb); a reacção ocorreu num volume total de 25 µL, com 1 µL do ADN/H₂O. **(d)** Reacção em *singleplex* com os *primers* F402/R619 (218 pb); a reacção ocorreu num volume total de 25 µL.

2.4.2. Sequenciação das amostras amplificadas em *singleplex*

Os produtos da PCR das amostras 109, 115, 117, 123, 124 e C+ amplificadas em reacções *singleplex* foram enviados para sequenciação para a empresa MacroGen, posteriormente à realização da purificação dos produtos da PCR pelo método ExoSAP-IT.

Da sequenciação do ADN das amostras de Goa obtiveram-se apenas electroforetogramas de reduzida qualidade, principalmente devida à elevada sobreposição de picos, não se tendo conseguido obter qualquer informação adicional sobre a respectiva amostra de Goa. Neste caso os resultados não são apresentados uma vez que não foram considerados para a determinação das sequências das amostras. Apenas o electroforetograma da sequência da amostra **controlo positivo**, sequenciada com o *primer* F402, apresenta picos bem definidos, sendo por isso de boa qualidade.

3. Afiliação em haplogrupos

As sequências das amostras caracterizadas por electroforetogramas de melhor qualidade foram alinhadas com a sequência CRSr, determinando-se assim a sequência de cada amostra e o haplótipo que caracteriza cada uma, associando-se cada amostra ao respectivo haplogrupo mitocondrial.

Na tabela 11 estão representados os haplótipos e haplogrupos associados a cada amostra de Goa. Em 10 amostras (amostras nº: 107, 108, 109, 110, 112, 113, 114, 115, 116, 117) o haplótipo determinado associa-as ao haplogrupo H, encontrado em elevada frequência na população europeia. A amostra 123 foi associada ao haplogrupo T2b, encontrado com frequência na população europeia. Na amostra 124 não foi possível determinar o haplogrupo com ambos os métodos devido à possibilidade de associação a múltiplos haplogrupos. A amostra 125 foi associada ao haplogrupo R5a/R5a2, encontrado com frequência na população asiática, em especial nas populações indianas e do Sri Lanka.

Tabela 11 – Determinação do haplogrupo de cada amostra através da associação dos respectivos haplótipos aos haplogrupos, com a utilização de duas ferramentas, PhyloTree e mtManager.

Amostra	Haplótipo	Haplogrupo estimado		
		<i>PhyloTree</i>	<i>mtManager</i>	Região geográfica de maior frequência
107	16519C, 263G	H	H	Europa
108	199C, 263G, 309.2C, 311.1C	H	H	Europa
109	16434A, 263G, 16519C, 309.1C, 311.1C	H	H	Europa
110	73G, 16519C,	H	H	Europa
112	73G, 146C	H	H	Europa
113	263G	H	H	Europa
114	73G, 152C, 16519C	H	H	Europa
115	263G, 309.1C, 311.1C	H	H	Europa
116	207A, 263G, 309.1C, 311.1C	H	H	Europa
117	263G, 538G, 311.1C	H	H	Europa
123	263G, 16126C, 16294T, 16296T, 16304C, 309.1C, 311.1C	T2b	T2b	Europa
124	73G, 263G	indeterminado	indeterminado	indeterminado
125	16266T, 16304C, 16325C, 16356C	R5a2	R5a	Ásia

As amostras caracterizadas por haplótipos associados ao haplogrupo H, são todas cedidas por habitantes de Goa com nome e/ou apelido de origem portuguesa, à excepção da amostra 117 que pertence a um indivíduo com nome indiano. A amostra 123 com haplótipo associado também a um haplogrupo mais frequente na Europa, foi cedida por um indivíduo com nome e apelido indianos. Já a amostra 125, com haplótipo associado a um haplogrupo característico da Ásia foi cedida por um indivíduo de apelido de origem portuguesa.

IV – Discussão

(Página intencionalmente deixada em branco)

1. Influência de inibidores da PCR

Na comparação de 2 diferentes intervalos de tempo decorridos entre a extracção e a amplificação das amostras através da PCR observaram-se, na generalidade, melhores resultados com as amostras com um intervalo de extracção inferior a 6 meses, com menores percentagens de amostras sem amplificação e maiores percentagens de amostras onde ocorreu amplificação de todos os fragmentos alvo (tabela 8). Já na comparação individual as diferenças não são evidentes, tendo-se obtido resultados semelhantes (tabela 9). O armazenamento das amostras durante um período superior a um ano sem realização prévia da separação do ADN da solução de Chelex, pode ter influenciado negativamente a amplificação por PCR do ADN extraído. Isto porque com o passar do tempo pode ocorrer degradação do Chelex, havendo libertação dos inibidores da PCR previamente capturados ou inibição da PCR pelo Chelex através da sua ligação aos iões magnésio, necessários ao funcionamento da polimerase (Lincoln & Thomson, 1998; Willard *et al.*, 1998)

Na quantificação do ADN por rt-qPCR testaram-se 2 *kits* comercial constituídos por fluorocromos diferentes, tendo-se detectado a presença de inibição da reacção de PCR na reacção com o fluorocromo *SybrGreen* (figura 11), mas não detectada com o fluorocromo *EvaGreen*. Num estudo realizado por Al-Soud & Radstrom, 2000, foi demonstrado que a adição de amostras com capacidade inibitória, inclusive amostras sanguíneas, reduz o sinal fluorescente durante a amplificação destas amostras com *SybrGreen*, ocorrendo uma relação linear entre a diminuição da concentração do inibidor e o aumento do sinal fluorescente. Os inibidores reduzem o número de moléculas do fluorocromo disponíveis para a ligação ao ADN de cadeia dupla, quer pela interferência com o *SybrGreen* e/ou o ADN de cadeia dupla (Al-Soud & Radstrom, 2000). A maior estabilidade térmica e hidrolítica do fluorocromo *EvaGreen* comparativamente ao *SybrGreen*, já foi demonstrada, podendo contribuir para uma menor sensibilidade à influência de inibidores na amplificação com o fluorocromo *EvaGreen* (Deligeorgiev *et al.*, 2009). O facto de diferentes polimerases apresentarem diferentes sensibilidades a inibidores da PCR, descrito por Al-Soud & Radstrom, 1998 e por Panaccio & Lew, 1991, pode também justificar os diferentes resultados obtidos, uma vez que cada fluorocromo foi utilizado com uma polimerase diferente.

Os inibidores da PCR detectados estão relacionados com o tipo de material biológico utilizado, nomeadamente sangue. Neste tipo de material biológico os inibidores da PCR são principalmente compostos provenientes do processo de degradação da hemoglobina sanguínea, como o heme e a hemina, ocorrendo também inibição por outros componentes sanguíneos como a lactoferrina e imunoglobina G (Akane *et al.*, 1994; Al-Soud *et al.*, 2000; Al-Soud & Radstrom, 2001). Os mecanismos de inibição dependem do componente inibidor, que pode inibir a reacção por redução da actividade da ADN polimerase ou por interacção com o ADN alvo da amplificação. Nos casos da hemoglobina e lactoferrina esta inibição pode estar relacionada com a capacidade de libertação de iões ferro que por sua vez competem com iões magnésio, cofactores da polimerase, pelo local de reacção, diminuindo a actividade de enzima (Al-Soud & Radstrom, 2001). No caso da imunoglobina G o efeito inibitório deve-se à sua ligação ao ADN de cadeia simples tornando-o indisponível para amplificação (Al-Soud *et al.*, 2000).

2. Presença de contaminação em amostras degradadas

A presença de contaminação detectada na análise de algumas amostras por PCR em tempo real (resultados não apresentados) revela a elevada sensibilidade da PCR em tempo real para a detecção do ADN contaminante mesmo quando presente em reduzida quantidade. No caso de análise do ADN de amostras degradadas, o risco de contaminação aumenta, uma vez que o número de ciclos de amplificação realizados é elevado (superior a 40 ciclos) para que haja detecção do reduzido número de cópias do ADN (Andreasson *et al.*, 2002). A proximidade de valores de Cq de amostras degradadas e do ADN contaminante potencialmente presente, torna provável a quantificação simultânea do ADN alvo e ADN contaminante, tendo-se optado por excluir da análise várias amostras cujos valores de Cq se aproximavam dos valores de Cq dos controlos negativos.

Apesar da presença do ADN alvo em maior número relativamente ao ADN contaminante, o estado de degradação da molécula torna-a um mau substrato para a ADN polimerase que leva a cabo uma reacção de amplificação instável, contrariamente ao que acontece com o ADN contaminante, caracterizado por uma estrutura bem conservada que serve de melhor substrato à ADN polimerase (Pruvost & Geigl, 2004). Na reacção onde ambas as moléculas do ADN estão presentes ocorre portanto a competição

pela ADN polimerase, que por sua vez prefere o ADN contaminante, fazendo com que este contribua também para a reacção de amplificação (Pruvost & Geigl, 2004).

A possibilidade de existência do ADN contaminante nas amostras de Goa é levantada pela análise dos electroforetogramas obtidos na sequenciação do ADN. Na maioria dos electroforetogramas de sequenciação verifica-se levantamento da linha de base ao longo da sequência com elevada frequência de sobreposição de picos em várias posições da sequência. A presença de elevados níveis de sobreposição impossibilitou a utilização de várias sequências obtidas por sequenciação. A associação desta sobreposição a ADN contaminante, e não a heteroplasmia pontual, deve-se á elevada frequência de posições sobrepostas sendo que muitas delas por vezes concentram-se na mesma região da sequência, contrariamente ao que acontece com a heteroplasmia, onde estas posições são pouco frequentes (Fendt *et al.*, 2009).

A elevada influência do ADN contaminante na sequenciação das amostras, como já referido, é consequência do elevado grau de degradação das amostras e do reduzido número de cópias do ADN de fragmentos longos, uma vez que não é visível o mesmo grau de presença de sobreposição de sequências com o controlo positivo também alvo de sequenciação com todos os *primers*. Este ADN contaminante pode dever-se à transferência do ADN entre as amostras ainda no suporte de tecido de algodão, uma vez que estas permaneceram armazenadas sem separação física entre elas por vários anos, até serem separadas em envelopes individuais, o que eventualmente poderá ter possibilitado a transferência do ADN entre as amostras.

As amostras analisadas apresentam elevada sensibilidade à degradação, sendo por isso importante tomar todas as precauções possíveis para evitar o contributo do ADN contaminante para os resultados obtidos com os diferentes métodos analisados. Uma vez que a grande problemática da manipulação do ADNmt está relacionada com o controlo e monitorização da contaminação, são várias as recomendações para a diminuição deste risco, principalmente aquando da análise do ADNmt degradado. As práticas laboratoriais devem incluir: áreas laboratoriais de pré e pós-amplificação devem ser separadas fisicamente, as áreas de trabalho devem ser limpas antes e depois da utilização, devem ser tomadas medidas como utilização de batas e material descartável como luvas e máscara, utilização de pontas de pipetas resistentes a aerossóis, os materiais e reagentes devem ser expostos á luz UV (Bar *et al.*, 2000; Carracedo *et al.*, 2000). Para monitorização da contaminação devem ser realizados brancos de reagentes e controlos negativos (amostras com reagentes mas sem ADN alvo) são a única forma de

monotorização da contaminação. Contudo, devido à sensibilidade de detecção da análise do ADNmt, reduzidos níveis de contaminação por ADN exógeno podem ser observáveis, sendo toleráveis baixos níveis de contaminação uma vez que resultados fiáveis podem ser obtidos na presença de contaminação (Bar *et al.*, 2000; Carracedo *et al.*, 2000).

3. Determinação da degradação do ADN através da quantificação

A quantificação do ADN com o fluorímetro demonstrou a existência do ADN nas amostras passível de possibilitar a amplificação do ADN. As amostras de Goa apresentaram quantidades do ADN semelhantes, o que é indicativo da homogeneidade entre as amostras de Goa (tabela 4). Quando comparado o valor médio da concentração do ADN das amostras de Goa com a concentração da amostra controlo positivo denota-se a superioridade da concentração média das amostras de Goa, contrariando aquilo que seria esperado, uma vez que a amostra controlo positivo é uma amostra de extracção recente, verificando-se mesmo no respectivo gel de agarose (figura 15 e figura 16) uma maior intensidade de sinal na amostra controlo positivo comparativamente às amostras de Goa.

Estes valores reduzidos podem dever-se a incompatibilidades entre o método de extracção do ADN utilizado em ambos os casos, extracção por Chelex, e o *kit* de quantificação utilizado. Isto porque a extracção do ADN pelo método de Chelex, devido á alcalinidade das soluções utilizadas provoca a desnaturação do ADN, obtendo-se no final do processo de extracção, ADN de cadeia simples.

Já o *kit* de quantificação utilizado é altamente selectivo para ADN de cadeia dupla, possuindo reduzidos níveis de detecção para ADN de cadeia simples. Tendo sido demonstrado que na presença de uma mistura do ADN de cadeia simples e ADN de cadeia dupla em partes iguais, a concentração do ADN detectada é apenas 10% superior aquando a detecção do ADN de cadeia dupla sozinho (Life Technologies, 2010). A metodologia utilizada para quantificação não é por este motivo adequada à análise do ADN extraído por Chelex, uma vez que não permite quantificar o ADN total.

A obtenção do ADN de cadeia simples limita a sua utilização, não sendo adequado para metodologias que dependam da presença do ADN de cadeia dupla, como análise por RFLP e quantificação por corantes intercalantes como o brometo de etídio,

sendo necessária a amplificação por PCR para a análise do ADN extraído por Chelex (Lincoln & Thomson, 1998).

De forma a obter-se a quantidade exacta do ADN extraído por Chelex procedeu-se à quantificação do ADN simultaneamente à sua amplificação por rt-qPCR.

Para a quantificação do ADN por PCR em tempo real, a escolha do fluorocromo recaiu sobre o *EvaGreen*, que não demonstrou influência de inibidores da PCR, tendo também apresentado maiores unidades de fluorescência mas também superiores valores de Cq quando comparado com o *SybrGreen*. Foram as duas primeiras, as características responsáveis pela escolha do *EvaGreen* para a quantificação do ADN, uma vez que já foi descrito por Mao *et al.*, 2007, que a diferença de valores Cq depende da sequência alvo, não havendo uma regra, sendo que o *SybrGreen* pode apresentar menores ou maiores valores de Cq dependendo da sequência alvo, não tendo por isso este parâmetro tido grande relevância para a escolha do fluorocromo.

Como resultados da quantificação por PCR em tempo real, à excepção da amostra 3, que apresenta 1.123 cópias iniciais do ADN, as amostras de Goa quantificadas apresentam valores bastante reduzidos do número de cópias iniciais do ADN, entre 232 e 93 cópias, sendo que em 3 amostras o número de cópias é tão reduzido que não foi possível a quantificação por não se localizar na curva padrão estimada. Estes resultados são indicativos do elevado grau de degradação das amostras estudadas, assemelhando-se aos resultados obtidos por Alonso *et al.*, 2004, entre 500 a 50 cópias iniciais do ADN, aquando da quantificação de um fragmento de 113 pb da região HVRI do ADNmt de amostras degradadas.

A degradação do ADN presente nas amostras pode estar relacionada com vários processos de degradação, como enzimáticos, oxidativos, hidrolíticos, entre outros. No caso destas amostras em sangue seco, a presença de hemoglobina é um factor que pode contribuir para o processo de degradação, isto porque a hemoglobina no estado seco, por exemplo em mancha de sangue, sofre alterações oxidativas e liberta catiões de ferro, criando a oportunidade para que ocorram danos na estrutura do ADN induzidos pelo radical hidroxilo (Marrone & Ballantyne, 2009).

4. Aplicação da estratégia multiplex

A análise do ADNmt consiste usualmente na amplificação e sequenciação de fragmentos longos do ADNmt, da totalidade da região não codificante do ADNmt, ou apenas das regiões hipervariáveis HVRI e HVRII. Contudo, quando as amostras a

analisar se encontram altamente degradadas não é possível a obtenção de fragmentos longos, recorrendo-se por isso a estratégias que dividem estes fragmentos em fragmentos de menor tamanho. A divisão da região não codificante do ADNmt em vários fragmentos torna a análise mais trabalhosa e dispendiosa, sendo por isso importante encontrar uma estratégia viável para obtenção de resultados mas também de fácil execução e custo reduzido.

A estratégia utilizada neste trabalho foi desenvolvida por Berger & Parson, 2009, como adaptação de uma estratégia já existente (Eichmann & Parson, 2008), com a diminuição do número de fragmentos alvo de 10 para 5 e com o aumento do tamanho dos amplicões de 144 a 237 pb para 288 a 444 pb, procurando-se assim amplificar também toda a região não codificante do ADNmt mas com menor complexidade, trabalho e custos associados.

A amplificação por PCR da região codificante em *multiplex* de fragmentos de tamanhos entre 288 e 444 pb evidenciou o elevado grau de degradação das amostras estudadas (tabela 6). Apesar de ter ocorrido amplificação na maioria das amostras estudadas, não foi possível amplificar todos os fragmentos alvo. A percentagem de sucesso de amplificação dos 5 fragmentos alvo foi bastante reduzida, de apenas 10% o equivalente a 13 amostras em 127 amostras estudadas. Em nenhuma das 13 amostras sequenciadas foi possível obter a sequência completa da região não codificante do ADNmt (tabela 10).

Os primeiros 20 pb da HVRI não foram obtidos em nenhum dos casos, o que indica a necessidade de alteração do *primer* F15989 para outro *primer* com local de ligação mais afastado do início da região HVRI, pois com a utilização deste *primer* a porção inicial desta região só poderia ser obtida com sequenciação com o *primer* reverso correspondente, R16433. A região HVRI revelou-se a região mais problemática para obtenção da sequência, com o amplicão de 444 pb a revelar-se bastante problemático, pois quando ocorreu amplificação deste fragmento, é notória a sua menor intensidade no gel de agarose comparativamente aos outros fragmentos visualizados (figura 16 e figura 18). A divisão deste amplicão em vários amplicões de menores tamanhos seria uma medida que poderia aumentar o sucesso de amplificação da região alvo, contudo quando tal foi efectuado através da adaptação para *singleplex* da estratégia de Eichmann & Parson, 2008, os fragmentos obtidos apresentaram baixa intensidade no gel de agarose (figura 19) e não possibilitaram a obtenção da sequência por sequenciação.

Em ambas as estratégias de amplificação utilizadas a dificuldade de obtenção dos fragmentos alvo na amplificação é evidenciada pela sequenciação, onde em alguns

casos, fragmentos amplificados por PCR não resultaram em sequências aquando da sua sequenciação. A não obtenção de sequências reflecte a reduzida quantidade do ADN presente nas amostras analisadas, que apesar de ter ocorrido amplificação do ADN, a quantidade resultante não é suficiente para a obtenção de resultados na sequenciação.

Em algumas amostras verificou-se a amplificação de um fragmento não alvo de tamanho superior aos fragmentos definidos como alvo de amplificação, o que é indicativo do menor grau de degradação destas amostras. Este fragmento forma-se possivelmente pela conjugação espontânea de *primers* presentes num determinado *multiplex* possibilitada pela presença de moléculas do ADN de maior tamanho nas amostras de Goa, tendo também ocorrido na amostra controlo positivo, o que sustenta esta hipótese, uma vez que esta amostra apresenta ADN não degradado. No caso do *multiplex* I ocorreu a amplificação de um fragmento de 760 pb resultante da conjugação do *primer* F15989 com o *primer* R180, evidenciado nos géis de agarose (figura 16) e confirmado pela sequenciação automática da amostra 109 onde o fragmento resultante da sequenciação com o *primer* F15989 consiste numa sequência de aproximadamente 733 pb que quando alinhada com a CRSr equivale ao intervalo das posições 16039 pb a 161 pb (Anexo 1). Já no *multiplex* II ocorreu amplificação de um fragmento de 832 pb, visível nos géis de agarose (figura 15), resultante da conjugação do *primer* F16197 com o *primer* R460, não tendo contudo aparecido o fragmento longo na sequenciação, o que poderá dever-se ao menor número de moléculas existentes, uma vez que no respectivo gel de agarose a intensidade da banda é reduzida.

Segundo Gabriel *et al.*, 2001 a amplificação de fragmentos do ADN curtos em amostras altamente degradadas apresenta benefícios como aumento no rendimento e especificidade em amostras altamente degradadas. Isto porque moléculas de menor tamanho estão presentes em maior abundância, o que permite uma aumento da produção das sequências do ADN alvo e uma diminuição na quantidade de produtos não-específicos.

As estratégias utilizadas, descritas como adequadas a amostras altamente degradadas e amostras mediantemente degradadas, na análise das amostras de Goa foram executadas com taxa de sucesso bastante reduzida, reflectindo um grau de degradação bastante elevado. Estas problemáticas reflectiram-se num reduzido número de amostras onde a amplificação de toda a região controlo revelou-se possível, 13 amostras (10%) em 127 nas amostras com intervalo de extracção/amplificação superior a 1 ano e de 9 amostras (32%) de 28 amostras analisadas com intervalo de extracção/amplificação

inferior a 6 meses. Sendo sempre a amplificação de apenas alguns fragmentos alvo os resultados mais comuns, o que permitiria apenas a obtenção de perfis genéticos incompletos. No caso de amostras forenses, mesmo resultados de perfis incompletos são um importante contributo para a investigação, uma vez que podem fornecer informações relevantes sobre as amostras, sendo nestes casos importante a obtenção das sequências do ADNmt das amostras. Contudo, para o enquadramento dos perfis genéticos num contexto populacional, um perfil genético incompleto não é suficiente, sendo critério mínimo o conhecimento das sequências das regiões HVRI e HVRII, com preferencialmente toda a região não codificante conhecida, não se tornando viável a obtenção de perfis genéticos incompletos ou de um reduzido número de amostras para este fim.

5. Diversidade genética da população de Goa

A diversidade haplotípica mitocondrial presente nas 13 amostras de Goa foi determinada com recurso à sequenciação do ADNmt. Contudo, devido à impossibilidade de obtenção da sequência completa da região não codificante, estes haplogrupos não devem ser encarados como definitivos, uma vez que são várias as regiões da sequência do ADNmt em falta, existindo por isso mutações desconhecidas que podem alterar o haplogrupo aqui determinado e associar as amostras a outro haplogrupo.

Neste estudo não foram observados os haplogrupos mais frequentemente observados na população indiana, nomeadamente o haplogrupo M encontrado em 60% a 64% da população e o haplogrupo U encontrado em 14% a 21% da população (Baig *et al.*, 2004; Kivisild *et al.*, 1999; Roychoudhury *et al.*, 2000; Thanseem *et al.*, 2006).

O haplogrupo com maior frequência neste estudo revelou ser o haplogrupo H, que é comumente observado em reduzida frequência em estudos já realizados, com uma frequência de 1,8% na Índia (Kivisild *et al.*, 1999). O haplogrupo H equivale a mais de 40% de toda a variação mitocondrial europeia, observando-se uma tendência para a sua diminuição à medida que as populações se localizam mais próximo da Ásia, onde é encontrado em 11% da população (Loogvali *et al.*, 2004).

O haplogrupo R5 associado a uma amostra, foi observado numa frequência de 10,66% no estudo de Thanseem *et al.*, 2006. Já o haplogrupo T, com o subhaplogrupo T2 associado a uma amostra, foi observado na população indiana com uma frequência de 0,86%, no estudo de Thanseem *et al.*, 2006 e de 1,8% no estudo de Kivisild *et al.*, 1999.

Na amostra da população de Goa analisada é evidente a elevada frequência do haplogrupo H, contrariamente ao documentado para o restante território indiano, onde a frequência deste haplogrupo é bastante reduzida. Contudo, uma avaliação ao nível da genética populacional, com as amostras disponíveis não apresenta grande fiabilidade, pois os haplogrupos são determinados de acordo com um conjunto de mutações em toda a região não codificante, sendo por isso importante o conhecimento da sequência da totalidade desta região ou de pelo menos das regiões hipervariáveis.

(Página intencionalmente deixada em branco)

V – Conclusão

(Página intencionalmente deixada em branco)

A sobrevivência do ANDmt por longos períodos de tempo em amostras não acondicionadas, como acontece em amostras forenses, torna-o o único recurso para obtenção de informação sobre a amostra pois nestes casos o ADN nuclear é escasso ou inexistente. Contudo com o decorrer do tempo as moléculas do ADNmt presentes na amostra sofrem processos de degradação, caracterizados por danos na sua estrutura, que têm como consequência a fragmentação da molécula do ADN em vários fragmentos de reduzidos tamanhos. Tal aconteceu com as amostras estudadas equiparando-as a amostras forenses com elevado grau de degradação, demonstrado por várias técnicas utilizadas. A elevada sensibilidade da PCR em tempo real tornou possível a detecção de reduzidas quantidades do ADN, revelando-se uma técnica sendo bastante útil para a determinação do grau de degradação de amostras degradadas.

O estado de degradação das amostras de Goa estudadas dificulta a realização de uma análise no âmbito da genética populacional, trazendo a necessidade de obtenção de novas amostras atendendo à relevância da população alvo deste trabalho. Assim, a obtenção de amostras novas do ADN da população de Goa seria uma oportunidade de concretização do estudo populacional de forma a determinar a estrutura genética que define os habitantes de Goa.

Neste trabalho foram analisadas 135 amostras degradadas de habitantes da população indiana de Goa, com 13 amostras sequenciadas por sequenciação automática. A aplicação das estratégias utilizadas para amplificação do ADNmt, nomeadamente a divisão da região não codificante em vários fragmentos alvo, revelou-se útil apenas na obtenção de perfis genéticos incompletos por sequenciação, o que permitiu a estimação de haplogrupos mitocondriais provisórios de um reduzido número de amostras da população de Goa.

Os haplogrupos observados nas amostras de Goa resumem-se ao haplogrupo H, o mais frequente, e aos haplogrupos T2b e R5a2 com a mesma frequência. O facto da frequência do haplogrupo H ser bastante elevada sugere a possibilidade de ter ocorrido influência portuguesa na população de Goa, o que evidencia a importância do estudo desta população indiana, uma vez que as interações com diferentes populações de origem europeia, sobretudo com a portuguesa, podem ter deixado marcas no genoma da população de Goa.

(Página intencionalmente deixada em branco)

VI – Referências Bibliográficas

(Página intencionalmente deixada em branco)

- Akane, A., K. Matsubara, H. Nakamura, S. Takahashi and K. Kimura, 1994 Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *J Forensic Sci* **39**: 362-372.
- Al-Soud, W. A., L. J. Jonsson and P. Radstrom, 2000 Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. *J Clin Microbiol* **38**: 345-350.
- Al-Soud, W. A., and P. Radstrom, 1998 Capacity of nine thermostable DNA polymerases To mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl Environ Microbiol* **64**: 3748-3753.
- Al-Soud, W. A., and P. Radstrom, 2000 Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces, and meat. *J Clin Microbiol* **38**: 4463-4470.
- Al-Soud, W. A., and P. Radstrom, 2001 Purification and characterization of PCR-inhibitory components in blood cells. *J Clin Microbiol* **39**: 485.
- Alaeddini, R., S. J. Walsh and A. Abbas, 2010 Forensic implications of genetic analyses from degraded DNA--A review. *Forensic Sci Int Genet* **4**: 148-157.
- Alonso, A., P. Martin, C. Albarran, P. Garcia, O. Garcia *et al.*, 2004 Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci Int* **139**: 141-149.
- Anderson, S., A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457-465.
- Andreasson, H., U. Gyllenstein and M. Allen, 2002 Real-time DNA quantification of nuclear and mitochondrial DNA in forensic analysis. *Biotechniques* **33**: 402-411.
- Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull *et al.*, 1999 Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetic* **23**: 147-147.
- Azevedo, C. (Editor), 2005 *Biologia Celular e Molecular*. Lidel.
- Baig, M., A. Khan and K. Kulkarni, 2004 Mitochondrial DNA diversity in tribal and caste groups of Maharashtra (India) and its implication on their genetic origins. *Ann Hum Genet* **68**: 453-460.
- Bamshad, M., T. Kivisild, W. S. Watkins, M. E. Dixon, C. E. Ricker *et al.*, 2001 Genetic evidence on the origins of Indian caste populations. *Genome Res* **11**: 994.
- Bar, W., B. Brinkmann, B. Budowle, A. Carracedo, P. Gill *et al.*, 2000 DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Int J Legal Med* **113**: 193-196.
- Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton *et al.*, 2008 Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**: 53-59.
- Berger, C., and W. Parson, 2009 Mini-midi-mito: adapting the amplification and sequencing strategy of mtDNA to the degradation state of crime scene samples. *Forensic Sci Int Genet* **3**: 149-153.
- Bertino, A. J., and P. N. Bertino, 2008 *Forensic Science: Fundamentals and Investigations*. South-Western Educational & Professional.
- Bhargava, S. C. B. G. K., 2006 *Land and people of Indian states and union territories: in 36 volumes*. Goa. Kalpaz publ.
- Bini, C., S. Ceccardi, D. Luiselli, G. Ferri, S. Pelotti *et al.*, 2003 Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). *Forensic Sci Int* **135**: 48-52.
- Bonatto, S. L., and F. M. Salzano, 1997 A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proc Natl Acad Sci U S A* **94**: 1866-1871.
- Borges, C. J., and H. Feldmann, 1997 *Goa and Portugal: their cultural links*. Concept Pub. Co.
- Brandon, M. C., M. T. Lott, K. C. Nguyen, S. Spolim, S. B. Navathe *et al.*, 2005 MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res* **33**: D611.
- Brandstatter, A., H. Niederstatter, M. Pavlic, P. Grubwieser and W. Parson, 2007 Generating population data for the EMPOP database - an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. *Forensic Sci Int* **166**: 164-175.
- Brown, M. D., S. H. Hosseini, A. Torroni, H. J. Bandelt, J. C. Allen *et al.*, 1998 mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* **63**: 1852-1861.
- Brown, W. M., 1980 Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci U S A* **77**: 3605-3609.
- Brown, W. M., M. George, Jr. and A. C. Wilson, 1979 Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* **76**: 1967-1971.
- Butler, J. M., 2009 *Fundamentals of Forensic DNA Typing*. Academic Press/Elsevier.
- Butler, J. M., 2011 *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Science.

- Calafell, F., P. Underhill, A. Tolun, D. Angelicheva and L. Kalaydjieva, 1996 From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* **60**: 35-49.
- Cann, R. L., M. Stoneking and A. C. Wilson, 1987 Mitochondrial DNA and human evolution. *Nature* **325**: 31-36.
- Cardoso, S., M. T. Zarrabeitia, L. Valverde, A. Odriozola, M. A. Alfonso-Sanchez *et al.*, 2010 Variability of the entire mitochondrial DNA control region in a human isolate from the Pas Valley (northern Spain). *J Forensic Sci* **55**: 1196-1201.
- Carracedo, A., W. Bär, P. Lincoln, W. Mayr, N. Morling *et al.*, 2000 DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* **110**: 79-85.
- Cavalli-Sforza, L. L., and M. W. Feldman, 2003 The application of molecular genetic approaches to the study of human evolution. *Nat Genet* **33 Suppl**: 266-275.
- Cavalli-Sforza, L. L., P. Menozzi and A. Piazza, 1994 *The history and geography of human genes*. Princeton University Press.
- Chandramouli, C., 2011 Census of India 2011, pp. in *Provisional Population Totals Paper 1 of 2011 India Series I*. Office of the Registrar General & Census Commissioner, India.
- Chary, M. T., 2009 *India: Nation on the Move: An Overview of India's People, Culture, History, Economy, It Industry, & More*. iUniverse.
- Clayton, D. A., J. N. Doda and E. C. Friedberg, 1974 The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. *Proc Natl Acad Sci U S A* **71**: 2777-2781.
- Coble, M. D., O. M. Loreille, M. J. Wadhams, S. M. Edson, K. Maynard *et al.*, 2009 Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* **4**: e4838.
- Cordaux, R., N. Saha, G. R. Bentley, R. Aunger, S. Sirajuddin *et al.*, 2003 Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* **11**: 253-264.
- Deligeorgiev, T. G., S. Kaloyanova and J. J. Vaquero, 2009 Intercalating Cyanine Dyes for Nucleic Acid Detection. *Recent Pat. Mater. Sci* **2**: 1-26.
- Educational, E. B., and K. Pletcher, 2010 *The Geography of India: Sacred and Historic Places*. Britannica Educational Pub. in association with Rosen Education Services.
- Edwin, D., H. Vishwanathan, S. Roy, M. Usha Rani and P. P. Majumder, 2002 Mitochondrial DNA diversity among five tribal populations of southern India. *Curr Sci India* **83**: 158-162.
- Eichmann, C., and W. Parson, 2008 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. *Int J Legal Med* **122**: 385-388.
- Fairchild, C. C., 2007 *Women in early modern Europe, 1500-1700*. Pearson/Longman.
- Fendt, L., B. Zimmermann, M. Daniaux and W. Parson, 2009 Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. *BMC genomics* **10**: 139.
- Forster, L., P. Forster, S. M. Gurney, M. Spencer, C. Huang *et al.*, 2010 Evaluating length heteroplasmy in the human mitochondrial DNA control region. *Int J Legal Med* **124**: 133-142.
- Forster, P., A. Torroni, C. Renfrew and A. Rohl, 2001 Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Mol Biol Evol* **18**: 1864-1881.
- Fridman, C., and R. Gonzalez, 2009 HVIII discrimination power to distinguish HVI and HVII common sequences. *Forensic Sci Int Genet* **2**: 320-321.
- Gabriel, M. N., E. F. Huffine, J. H. Ryan, M. M. Holland and T. J. Parsons, 2001 Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J Forensic Sci* **46**: 247-253.
- Gates, K. S., 2009 An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem Res Toxicol* **22**: 1747-1760.
- Giles, R. E., H. Blanc, H. M. Cann and D. C. Wallace, 1980 Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **77**: 6715-6719.
- Gill, P., 2001 Application of low copy number DNA profiling. *Croat Med J* **42**: 229-232.
- Gill, P., P. L. Ivanov, C. Kimpton, R. Piercy, N. Benson *et al.*, 1994 Identification of the remains of the Romanov family by DNA analysis. *Nature genetic* **6**: 130-135.
- Ginther, C., L. Issel-Tarver and M. C. King, 1992 Identifying individuals by sequencing mitochondrial DNA from teeth. *Nature genetic* **2**: 135-138.
- Goebel, T., M. R. Waters and D. H. O'Rourke, 2008 The late Pleistocene dispersal of modern humans in the Americas. *Science* **319**: 1497-1502.
- Goodwin, W., A. Linacre and S. Hadi, 2011 *An Introduction to Forensic Genetics*. John Wiley & Sons.
- Gracias, F. S., 1996 *Kaleidoscope of women in Goa, 1510-1961*. Concept Pub. Co.

- Greenberg, B. D., J. E. Newbold and A. Sugino, 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* **21**: 33-49.
- Grignani, P., G. Peloso, P. Fattorini and C. Previdere, 2009 A mini-primer set covering the mtDNA hypervariable regions for the genetic typing of old skeletal remains. *Forensic Sci Int Genet* **2**: 265-266.
- Grossman, L. I., R. Watson and J. Vinograd, 1973 The presence of ribonucleotides in mature closed-circular mitochondrial DNA. *Proc Natl Acad Sci U S A* **70**: 3339-3343.
- Hagelberg, E., 2003 Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. *Trends Genet* **19**: 84-90.
- Hall, A., and J. Ballantyne, 2004 Characterization of UVC-induced DNA damage in bloodstains: forensic implications. *Anal Bioanal Chem* **380**: 72-83.
- Hall, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, pp. 95-98.
- Hirszfeld, L., and H. Hirszfeld, 1919 Essai d'application des méthodes sérologiques au problème des races. *Anthropologie* **29**: 505-537.
- Howard, R., V. Encheva, J. Thomson, K. Bache, Y. T. Chan *et al.*, 2011 Comparative analysis of human mitochondrial DNA from World War I bone samples by DNA sequencing and ESI-TOF mass spectrometry. *Forensic Sci Int Genet*.
- Howell, N., J. L. Elson, C. Howell and D. M. Turnbull, 2007 Relative rates of evolution in the coding and control regions of African mtDNAs. *Mol Biol Evol* **24**: 2213.
- Human genome at ten: The sequence explosion. 2010 *Nature* **464**: 670-671.
- Human genome: Genomes by the thousand. 2010 *Nature* **467**: 1026-1027.
- Hutchison, C. A., 3rd, J. E. Newbold, S. S. Potter and M. H. Edgell, 1974 Maternal inheritance of mammalian mitochondrial DNA. *Nature* **251**: 536-538.
- Information, I. M. o., and Broadcasting, 2010 *India: a reference annual*. Publications Division, Ministry of Information and Broadcasting.
- Ingman, M., and U. Gyllensten, 2006 mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res* **34**: D749.
- Ingman, M., H. Kaessmann, S. Paabo and U. Gyllensten, 2000 Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708-713.
- Irwin, J. A., W. Parson, M. D. Coble and R. S. Just, 2010 mtGenome reference population databases and the future of forensic mtDNA analysis. *Forensic Sci Int Genet*.
- Ivanov, P. L., M. J. Wadhams, R. K. Roby, M. M. Holland, V. W. Weedn *et al.*, 1996 Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature genetic* **12**: 417-420.
- Just, R. S., O. M. Loreille, J. E. Molto, D. A. Merriwether, S. R. Woodward *et al.*, 2011 Titanic's unknown child: the critical role of the mitochondrial DNA coding region in a re-identification effort. *Forensic Sci Int Genet* **5**: 231-235.
- Kaneda, H., J. Hayashi, S. Takahama, C. Taya, K. F. Lindahl *et al.*, 1995 Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. *PNAS* **92**: 4542.
- Keyer, K., and J. A. Imlay, 1996 Superoxide accelerates DNA damage by elevating free-iron levels. *PNAS* **93**: 13635.
- Kivisild, T., M. J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu *et al.*, 1999 Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* **9**: 1331-1334.
- Knudson, K. J., and C. M. Stojanowski, 2008 New directions in bioarchaeology: recent contributions to the study of human social identities. *J Archaeol Res* **16**: 397-432.
- Kong, Q. P., H. J. Bandelt, M. Zhao, Y. P. Zhang and Y. G. Yao, 2010 Reply to van Oven: suggestions and caveats for naming mtDNA haplogroup. *PNAS* **107**: E40.
- Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebraldise, W. S. Kunz *et al.*, 2004 Recombination of human mitochondrial DNA. *Science* **304**: 981.
- Lamers, R., S. Hayter and C. D. Matheson, 2009 Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. *J Mol Evol* **68**: 40-55.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody *et al.*, 2001 Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Landsteiner, K., 1900 Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Centralblatt für Bacteriologie* **27**: 357-362.

- Lee, H. Y., I. Song, E. Ha, S.-B. Cho, W. I. Yang *et al.*, 2008 mtDNAManager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences. *BMC bioinformatics* **9**: 483.
- Legros, F., F. Malka, P. Frachon, A. Lombes and M. Rojo, 2004 Organization and dynamics of human mitochondrial DNA. *J Cell Sci* **117**: 2653-2662.
- Levy, S., G. Sutton, P. C. Ng, L. Feuk, A. L. Halpern *et al.*, 2007 The diploid genome sequence of an individual human. *PLoS biology* **5**: e254.
- Life Technologies, 2010 Single-stranded DNA in the Qubit™ dsDNA Assays for the Qubit® 2.0 Fluorometer, pp.
- Lincoln, P. J., and J. Thomson, 1998 *Forensic DNA profiling protocols*. Humana Press.
- Lodish, H. F., 2008 *Molecular cell biology*. W.H. Freeman.
- Loogvali, E. L., U. Roostalu, B. A. Malyarchuk, M. V. Derenko, T. Kivisild *et al.*, 2004 Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* **21**: 2012-2021.
- Lutz-Bonengel, S., T. Sanger, W. Parson, H. Muller, J. W. Ellwart *et al.*, 2008 Single lymphocytes from two healthy individuals with mitochondrial point heteroplasmy are mainly homoplasmic. *Int J Legal Med* **122**: 189-197.
- Lutz, S., H. J. Weisser, J. Heizmann and S. Pollak, 1997 A third hypervariable region in the human mitochondrial D-loop. *Hum Genet* **101**: 384.
- Lutz, S., H. J. Weisser, J. Heizmann and S. Pollak, 1998 Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* **111**: 67-77.
- Lutz, S., H. Wittig, H. J. Weisser, J. Heizmann, A. Junge *et al.*, 2000 Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Sci Int* **113**: 97-101.
- Mabuchi, T., R. Susukida, A. Kido and M. Oya, 2007 Typing the 1.1 kb control region of human mitochondrial DNA in Japanese individuals. *J Forensic Sci* **52**: 355-363.
- Macaulay, V., C. Hill, A. Achilli, C. Rengo, D. Clarke *et al.*, 2005 Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* **308**: 1034-1036.
- Mao, F., W. Y. Leung and X. Xin, 2007 Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC biotechnology* **7**: 76.
- Marrone, A., and J. Ballantyne, 2009 Changes in dry state hemoglobin over time do not increase the potential for oxidative DNA damage in dried blood. *PLoS One* **4**: e5110.
- Marrone, A., and J. Ballantyne, 2010 Hydrolysis of DNA and its molecular components in the dry state. *Forensic Sci Int Genet* **4**: 168-177.
- McEvoy, B. P., J. E. Powell, M. E. Goddard and P. M. Visscher, 2011 Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res* **21**: 821-829.
- Melton, T., 2004 Mitochondrial DNA heteroplasmy. *Forensic Sci Rev* **16**: 1-20.
- Messina, F., G. Scorrano, C. M. Labarga, M. F. Rolfo and O. Rickards, 2010 Mitochondrial DNA variation in an isolated area of Central Italy. *Ann Hum Biol* **37**: 385-402.
- Meyer, S., G. Weiss and A. Von Haeseler, 1999 Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* **152**: 1103.
- Mills, R. E., W. S. Pittard, J. M. Mullaney, U. Farooq, T. H. Creasy *et al.*, 2011 Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res* **21**: 830-839.
- Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn *et al.*, 1986 Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**: 263-273.
- Mullis, K. B., and F. A. Faloona, 1987 Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Method Enzymol* **155**: 335-350.
- Nag, P., and S. Sengupta, 1992 *Geography of India*. Concept Pub. Co.
- Nelson, D. L., A. L. Lehninger and M. M. Cox, 2008 *Lehninger principles of biochemistry*. W. H. Freeman.
- Nunez, C., M. Baeta, C. Sosa, Y. Casalod, J. Ge *et al.*, 2010 Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. *Am J Phys Anthropol* **143**: 591-600.
- Paabo, S., H. Poinar, D. Serre, V. Jaenicke-Despres, J. Hebler *et al.*, 2004 Genetic analyses from ancient DNA. *Annu Rev Genet* **38**: 645-679.
- Pakendorf, B., and M. Stoneking, 2005 Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* **6**: 165-183.

- Panaccio, M., and A. Lew, 1991 PCR based diagnosis in the presence of 8% (v/v) blood. *Nucleic Acids Res* **19**: 1151.
- Papiha, S. S., R. Deka and R. Chakraborty, 1999 Genomic diversity: applications in human population genetics, pp., edited by K. A. P. PUBLISHERS, New York.
- Parsons, T. J., and M. D. Coble, 2001 Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croat Med J* **42**: 304-309.
- Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Alliston-Greiner *et al.*, 1997 A high observed substitution rate in the human mitochondrial DNA control region. *Nature genetic* **15**: 363-368.
- Passarge, E., 2007 *Color atlas of genetics*. Thieme.
- Pauling, L., H. A. Itano and *et al.*, 1949 Sickle cell anemia a molecular disease. *Science* **110**: 543-548.
- Pruvost, M., and E. M. Geigl, 2004 Real-time quantitative PCR to assess the authenticity of ancient DNA amplification. *J Archaeol Res Sci* **31**: 1191-1197.
- Quintana-Murci, L., O. Semino, H. J. Bandelt, G. Passarino, K. McElreavey *et al.*, 1999 Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* **23**: 437-441.
- Raven, P. H., and G. B. Johnson, 1999 *Biology*. WCB/McGraw-Hill.
- Reidla, M., T. Kivisild, E. Metspalu, K. Kaldma, K. Tambets *et al.*, 2003 Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet* **73**: 1178-1190.
- Richards, M., H. Corte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots *et al.*, 1996 Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* **59**: 185.
- Richards, M., V. Macaulay, E. Hickey, E. Vega, B. Sykes *et al.*, 2000 Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**: 1251-1276.
- Richter, C., J. W. Park and B. N. Ames, 1988 Normal oxidative damage to mitochondrial and nuclear DNA is extensive. *Proc Natl Acad Sci U S A* **85**: 6465-6467.
- Risley, H., and W. Crooke, 1999 *The people of India*. Asian Educational Services.
- Robin, E. D., and R. Wong, 1988 Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol* **136**: 507-513.
- Roychoudhury, S., S. Roy, B. Dey, M. Chakraborty, M. Roy *et al.*, 2000 Fundamental genomic unity of ethnic India is revealed by analysis of mitochondrial DNA. *Curr Sci India* **79**: 1182-1192.
- Scarpulla, R. C., 2008 Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol Rev* **88**: 611-638.
- Schurr, T. G., and S. T. Sherry, 2004 Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* **16**: 420-439.
- Schwartz, M., and J. Vissing, 2002 Paternal inheritance of mitochondrial DNA. *N Engl J Med* **347**: 576-580.
- Shadel, G. S., and D. A. Clayton, 1997 Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem* **66**: 409-435.
- Shukla, A., 2009 *Population Genetics*. Discovery Publishing House Pvt. Ltd.
- Soares, P., L. Ermini, N. Thomson, M. Mormina, T. Rito *et al.*, 2009 Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* **84**: 740-759.
- Sondheimer, N., C. E. Glatz, J. E. Tirone, M. A. Deardorff, A. M. Krieger *et al.*, 2011 Neutral mitochondrial heteroplasmy and the influence of aging. *Hum Mol Gen* **20**: 1653-1659.
- Speicher, M. R., S. E. Antonarakis and A. G. Motulsky, 2010 *Vogel and Motulsky's Human Genetics: Problems and Approaches* Springer.
- Stoneking, M., 2000 Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet* **67**: 1029-1032.
- Stoneking, M., and F. Delfin, 2010 The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol* **20**: R188-193.
- Stubbe, H., 1972 *History of genetics: from prehistoric times to the rediscovery of Mendel's laws*. MIT Press.
- Sudoyo, H., P. T. Widodo, H. Suryadi, Y. S. Lie, D. Safari *et al.*, 2008 DNA analysis in perpetrator identification of terrorism-related disaster: Suicide bombing of the Australian Embassy in Jakarta 2004. *Forensic Sci Int Genet* **2**: 231-237.
- Tamm, E., T. Kivisild, M. Reidla, M. Metspalu, D. G. Smith *et al.*, 2007 Beringian standstill and spread of Native American founders. *PLoS One* **2**: e829.
- Thanseem, I., K. Thangaraj, G. Chaubey, V. Singh, L. Bhaskar *et al.*, 2006 Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC genetics* **7**: 42.
- Torroni, A., A. Achilli, V. Macaulay, M. Richards and H. J. Bandelt, 2006 Harvesting the fruit of the human mtDNA tree. *Trends Genet* **22**: 339-345.

- Underhill, P. A., and T. Kivisild, 2007 Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* **41**: 539-564.
- van Oven, M., and M. Kayser, 2009 Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**: E386-394.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural *et al.*, 2001 The sequence of the human genome. *Science* **291**: 1304.
- Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson, 1991 African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503-1507.
- Wai, T., D. Teoli and E. A. Shoubridge, 2008 The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat Genet* **40**: 1484-1488.
- Walsh, P. S., D. A. Metzger and R. Higuchi, 1991 Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506.
- Wang, J., W. Wang, R. Li, Y. Li, G. Tian *et al.*, 2008 The diploid genome sequence of an Asian individual. *Nature* **456**: 60-65.
- Warby, S. C., H. Visscher, J. A. Collins, C. N. Doty, C. Carter *et al.*, 2011 HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. *Eur J Hum Genet* **19**: 561-566.
- Wasnik, A. K., 2011 Census of India 2011, pp. in *Provisional Population Totals Paper 1 of 2011 Goa Series 31*. Office of the Registrar General & Census Commissioner.
- Watson, E., P. Forster, M. Richards and H. J. Bandelt, 1997 Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**: 691-704.
- Watson, J. D., and A. Berry, 2004 *DNA: the secret of life*. Arrow.
- Watson, J. D., and F. H. C. Crick, 1953 Molecular structure of nucleic acids. *Nature* **171**: 737-738.
- Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen *et al.*, 2008 The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**: 872-876.
- White, D., J. Wolff, M. Pierson and N. Gemmell, 2008 Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol* **17**: 4925.
- Willard, J. M., D. A. Lee and M. M. Holland, 1998 Recovery of DNA for PCR amplification from blood and forensic samples using a chelating resin, pp. 9-18 in *Forensic DNA Profiling Protocols*, edited by P. J. LINCOLN and J. THOMSON. Humana Press, Totowa, N.J.
- Xing, J., W. S. Watkins, Y. Hu, C. D. Huff, A. Sabo *et al.*, 2010 Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol* **11**: R113.
- Zhirkova, V., S. Fedorova, V. Akhmetova, L. Zhivotovsky and E. Khusnutdinova, 2011 Allelic polymorphism of six microsatellite DNA Loci in populations of Sakha (Yakutia). *Molecular Biology* **45**: 221-228.

Anexos

(Página intencionalmente deixada em branco)

Anexo 1: Alinhamento das sequências obtidas por sequenciação automática das amostras de Goa. CRSr indica a sequência referência de Cambridge revista, as restantes sequências são das amostras de Goa, com o número da amostra seguido pelo nome do *primer* utilizado para sequenciação. A fundo preto estão representadas as posições onde ocorreu mutação.

	10	20	30	40	50
CRSr	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
107_F109	-----	-----	-----	-----	-----
107_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
109_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
113_F109	-----	-----	-----	-----	-----
114_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----GT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	GATCACAGGT	CTATCACCCCT	ATTAACCACT	CACGGGAGCT	CTCCATGCAT
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	60	70	80	90	100
rCRS	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
107_F109	-----	-----	-----	-----	-----
107_F16450	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
109_F16450	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG

112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	TTGGTATTTT	CGTCTGGGGG	GTCTGCACGC	GATAGCATTG	CGAGACGCTG
113_F109	-----	-----	-----	-----	-----
114_F16450	TTGGTATTTT	CGTCTGGGGG	GTCTGCACGC	GATAGCATTG	CGAGACGCTG
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	TTGGTATTTT	CGTCTGGGGG	GTATGCACGC	GATAGCATTG	CGAGACGCTG
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	TTGGTATTTT	CGTCTGGGGG	GTCTGCACGC	GATAGCATTG	CGAGACGCTG
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	110	120	130	140	150
CRS _r	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
107_F109	-----	-----	-----	-----	-----
107_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
109_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCCATC
113_F109	-----	-----	-----	-----	-----
114_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	GAGCCGGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----

124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	GAGCCGAGC	ACCCTATGTC	GCAGTATCTG	TCTTTGATTC	CTGCCTCATC
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	160	170	180	190	200
CRS _r	CTATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
107_F109	---TTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
107_F16450	CTATTATTTA	TCGCACCTA-	-----	-----	-----
108_F109	--ATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACCA
109_F109	--ATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
109_F317	-----	-----	-----	-----	-----
109_F15989	CTATTATTTA	T-----	-----	-----	-----
109_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATATT	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATAT-	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATATT	-----	-----
113_F109	--ATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
114_F16450	CAATTATTTA	TCGCACCTAC	GTTCAATATT	-----	-----
115_F109	---TTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATATT	-----	-----
116_F109	CTATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
117_F109	CTATTATTTA	TCGCACCTAC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATATT	A-----	-----
123_F109	-----	-----AC	GTTCAATATT	ACAGGCGAAC	ATACTTACTA
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	--AGGCGAAC	ATACTTACTA
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	CTATTATTTA	TCGCACCTAC	GTTCAATA--	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	210	220	230	240	250
CRS _r	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
107_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
107_F16450	-----	-----	-----	-----	-----
108_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
109_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----

110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
114_F16450	-----	-----	-----	-----	-----
115_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
117_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	AAGTGTGTTA	ATTAATTAAT	GCTTGTAGGA	CATAATAATA	ACAATTGAAT
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	260	270	280	290	300
CRS _r	GTCTGCACAG	CCACTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
107_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
107_F16450	-----	-----	-----	-----	-----
108_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
109_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACC-
114_F16450	-----	-----	-----	-----	-----
115_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
117_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	GTCTGCACAG	CCGCTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----

123_F16197	-----	-----	-----	-----	-----
124_F109	GTCTGCACAG	CCCTTTTCCA	CACAGACATC	ATAACAAAAA	ATTTCCACCA
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	310	320	330	340	350
CRS _r	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
107_F109	AA-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
109_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
117_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	AACCCCCCCT	CCCCCGCTTC	TGGCCACAGC	ACTTAAACAC	ATCTCTGCCA
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	360	370	380	390	400
CRS _r	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
109_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
109_F317	-----	-----	-----G	CCTAACCAGA	TTTCAAATTT
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----

110_F317	-----	-----	-TAACACCAG	CCTAACCAGA	TTTCAAATTT
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
115_F317	-----	-----C	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
117_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
117_F317	-----	-----	-----	-----GA	TTTCAAATTT
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
123_F317	-----	---AAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	AACCCCAAAA	ACAAAGAACC	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	CTAACACCAG	CCTAACCAGA	TTTCAAATTT
125_F16197	-----	-----	-----	-----	-----

	410	420	430	440	450
CRSr	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
109_F109	TAT-----	-----	-----	-----	-----
109_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
115_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
117_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
117_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAA-----

123_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACA--
124_F317	-----G	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	TATCTTTTGG	CGGTATGCAC	TTTTAACAGT	CACCCCCCAA	CTAACACATT
125_F16197	-----	-----	-----	-----	-----

	460	470	480	490	500
CRSr	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	ATTTTCCCC-	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	ATT-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	C-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	ATTTTCCCCT	-----	-----	-----	-----
115_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	ATTTTCCCCT	-----	-----	-----	-----
117_F109	ATTTTCCCCT	-----	-----	-----	-----
117_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	ATTTTCCCCT	CCCACTCCCA	TACTACTAAT	CTCATCAATA	CAACCCCCGC
125_F16197	-----	-----	-----	-----	-----

	510	520	530	540	550
CRSr	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCAATA	CCCCGAACCA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----

109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCATA	CCCCGAACCA
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCATA	CCCCGAACCA
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCATA	CCCCGAACCA
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCATA	CCCCGAACCA
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	CCATCCTACC	CAGCACACAC	ACACCGCTGC	TAACCCCATA	CCCCGAACCA
125_F16197	-----	-----	-----	-----	-----

	560	570	580	590	600
CRS _r	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCTTA	CCTCCTCAAA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCTTA	CCTCCTCAAA
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCTTA	CCTCCTCAAA
117_F16197	-----	-----	-----	-----	-----

117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCT--	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCT--	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	ACCAAACCCC	AAAGACACCC	CCCACAGTTT	ATGTAGCT--	-----
125_F16197	-----	-----	-----	-----	-----
.... 					
610 620 630 640 650					
CRSr	GCAATACACT	GAAAAATGTTT	AGACGGGCTC	ACATCACCCC	ATAAACAAAT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----	-----
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	GCAATACACT	GAAAAATGTT-	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	GCAATACACT	GAAAAATGTT-	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

(intervalo 651 pb a 16000 pb não analisado)

	16010	16020	16030	16040	16050
CRS_r	ATTCTAATT	AAACTATTCT	CTGTTCTTTC	ATGGGGAAGC	AGATTGCGGT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	-----	-----	-----	-----GC	AGATTGCGGT
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----TGGGGAAGC	AGATTGCGGT	-----
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----TGGGGAAGC	AGATTGCGGT	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	16060	16070	16080	16090	16100
CRS_r	ACCACCCAAG	TATTGACTCA	CCCATCAACA	ACCGCTATGT	ATTTCGTACA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	ACCACCCAAG	TATTGACTCA	CCCATCAACA	ACCGCTATGT	ATTTCGTACA
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----

115_F317	-----	-----	-----	-----	-----
115_F15989	ACCACCCAAG	TATTGACTCA	CCCATCAACA	ACCGCTATGT	ATTTTCGTACA
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	ACCACCCAAG	TATTGACTCA	CCCATCAACA	ACCGCTATGT	ATTTTCGTACA
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	16110	16120	16130	16140	16150
CRS _r	TTACTGCCAG	CCACCATGAA	TATTGTACGG	TACCATAAAT	ACTTGACCAC
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	TTACTGCCAG	CCACCATGAA	TATTGTACGG	TACCATAAAT	ACTTGACCAC
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----G	TACCATAAAT	ACTTGACCAC
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	TTACTGCCAG	CCACCATGAA	TATTGTACGG	TACCATAAAT	ACTTGACCAC
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	TTACTGCCAG	CCACCATGAA	TATTG C ACGG	TACCATAAAT	ACTTGACCAC
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	16160	16170	16180	16190	16200
CRS_r	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----	-----
110_F16450	-----	-----	-----	-----	-----
112_F15989	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	CTGTAGTACA	TAAAAACCCA	ATCCACATCA	AAACCCCCTC	CCCATGCTTA
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	16210	16220	16230	16240	16250
CRS_r	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	-----	-----	-----	-----TCA	ACTGCAACTC
110_F16450	-----	-----	-----	-----	-----
112_F15989	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC
112_F16197	-----	-----	-----	-----	-----
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----

115_F317	-----	-----	-----	-----	-----
115_F15989	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC
115_F16197	-----	-----	-----	-----	-----
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	-----	-----	-----	-----CA	ACTGCAACTC
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	CAAGCAAGTA	CAGCAATCAA	CCCTCAACTA	TCACACATCA	ACTGCAACTC
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

	16260	16270	16280	16290	16300
CRS _r	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
110_F16450	-----	-----	-----	-----	-----
112_F15989	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
112_F16197	-----ACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
115_F16197	--AAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	CAAAGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCATCTTTAA
123_F16197	---AGCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCATCTTTAA
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	----GCCACC	CCTCACCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	----GCCACC	CCTCATCCAC	TAGGATACCA	ACAAACCTAC	CCACCCTTAA

	16310	16320	16330	16340	16350
CRS_r	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
110_F16450	-----	-----	-----	-----	-----
112_F15989	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
112_F16197	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
115_F16197	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	CAGC [■] ACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
123_F16197	CAGC [■] ACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	CAGTACATAG	TACATAAAGC	CATTTACCGT	ACATAGCACA	TTACAGTCAA
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	CAGC [■] ACATAG	TACATAAAGC	CATT [■] CACCGT	ACATAGCACA	TTACAGTCAA

	16360	16370	16380	16390	16400
CRS_r	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
110_F16450	-----	-----	-----	-----	-----
112_F15989	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCC-----
112_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----

115_F317	-----	-----	-----	-----	-----
115_F15989	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
115_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
123_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	ATCCCTTCTC	GTCCCCATGG	ATGACCCCCC	TCAGATAGGG	GTCCCTTGAC

	16410	16420	16430	16440	16450
CRS _r	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAAAGTGCT	ACTCTCCTCG
109_F16450	-----	-----	-----	-----	-----
110_F317	-----	-----	-----	-----	-----
110_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
110_F16450	-----	-----	-----	-----	-----
112_F15989	-----	-----	-----	-----	-----
112_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAA-----	-----
115_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	CACCATCCTC	CGTGAAATCA	ATA-----	-----	-----
123_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	CACCATCCTC	CGTGAAATCA	ATATCCCGCA	CAAGAGTGCT	ACTCTCCTCG

	16460	16470	16480	16490	16500
CRS_r	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-----	-----	-----	-----
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
109_F16450	-----	-----	-----	-----	-----T
110_F317	-----	-----	-----	-----	-----
110_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
110_F16450	-----	-----	-----	-----	-----T
112_F15989	-----	-----	-----	-----	-----
112_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
112_F16450	-----	-----	-----	-----	-----
113_F109	-----	-----	-----	-----	-----
114_F16450	-----	-----	-----	-----	-----
115_F109	-----	-----	-----	-----	-----
115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
115_F16450	-----	-----	-----	-----	-----
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGACAT
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGAC--
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCG----
124_F16450	-----	-----	-----	-----	-----
125_F317	-----	-----	-----	-----	-----
125_F16197	CTCCGGGCCC	ATAACACTTG	GGGGTAGCTA	AAGTGAAGTG	TATCCGA---

	16510	16520	16530	16540	16550
CRS_r	CTGGTTCCTA	CTTCAGGGTC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
107_F109	-----	-----	-----	-----	-----
107_F16450	-----	-TTCAGGGC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
108_F109	-----	-----	-----	-----	-----
109_F109	-----	-----	-----	-----	-----
109_F317	-----	-----	-----	-----	-----
109_F15989	CTGGTTCCTA	CTTCAGGGC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
109_F16450	CTGGTTCCTA	CTTCAGGGC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
110_F317	-----	-----	-----	-----	-----
110_F16197	CTGGTTCCT-	-----	-----	-----	-----
110_F16450	CTGGTTCCTA	CTTCAGGGC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
112_F15989	-----	-----	-----	-----	-----
112_F16197	CTGGTTCCT-	-----	-----	-----	-----
112_F16450	-----TA	CTTCAGGGTC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
113_F109	-----	-----	-----	-----	-----
114_F16450	---GTTTCCTA	CTTCAGGGC	ATAAAGCCTA	AATAGCCCAC	ACGTTCCCCT
115_F109	-----	-----	-----	-----	-----

115_F317	-----	-----	-----	-----	-----
115_F15989	-----	-----	-----	-----	-----
115_F16197	CTGGTTCCT-	-----	-----	-----	-----
115_F16450	-----	-----	----AGCCTA	AATAGCCCAC	ACGTTCCCCT
116_F109	-----	-----	-----	-----	-----
117_F109	-----	-----	-----	-----	-----
117_F317	-----	-----	-----	-----	-----
117_F16197	CTGGTTCCT-	-----	-----	-----	-----
117_F16450	-----	-----	-----	-----	-----
123_F109	-----	-----	-----	-----	-----
123_F317	-----	-----	-----	-----	-----
123_F15989	-----	-----	-----	-----	-----
123_F16197	-----	-----	-----	-----	-----
124_F109	-----	-----	-----	-----	-----
124_F317	-----	-----	-----	-----	-----
124_F16197	-----	-----	-----	-----	-----
124_F16450	-----	-----	-----	---AGCCCAC	ACGTTCCCCT
125_F317	-----	-----	-----	-----	-----
125_F16197	-----	-----	-----	-----	-----

....|....||....

16560

CRSr	TAAATAAGAC	ATCACGATG
107_F109	-----	-----
107_F16450	TAAATAAGAC	ATCACGATG
108_F109	-----	-----
109_F109	-----	-----
109_F317	-----	-----
109_F15989	TAAATAAGAC	ATCACGATG
109_F16450	TAAATAAGAC	ATCACGATG
110_F317	-----	-----
110_F16197	-----	-----
110_F16450	TAAATAAGAC	ATCACGATG
112_F15989	-----	-----
112_F16197	-----	-----
112_F16450	TAAATAAGAC	ATCACGATG
113_F109	-----	-----
114_F16450	TAAATAAGAC	ATCACGATG
115_F109	-----	-----
115_F317	-----	-----
115_F15989	-----	-----
115_F16197	-----	-----
115_F16450	TAAATAAGAC	ATCACGATG
116_F109	-----	-----
117_F109	-----	-----
117_F317	-----	-----
117_F16197	-----	-----
117_F16450	-----	-----
123_F109	-----	-----
123_F317	-----	-----
123_F15989	-----	-----
123_F16197	-----	-----
124_F109	-----	-----
124_F317	-----	-----
124_F16197	-----	-----
124_F16450	TAAATAAGAC	ATCACGATG
125_F317	-----	-----
125_F16197	-----	-----